JOURNAL OF MANAGEMENT AND ENGINEERING INTEGRATION

Scope: The Journal of Management and Engineering Integration (JMEI) is a double-blind refereed journal dedicated to exploring the nexus of management and engineering issues of the day. JMEI publishes two issues per year, one in the Summer and another in Winter. The Journal's scope is to provide a forum where engineering and management professionals can share and exchange their ideas for the collaboration and integration of Management and Engineering research and publications. The journal will aim on targeting publications and research that emphasizes the integrative nature of business, management, computers and engineering within a global context.

EDITORIAL REVIEW BOARD

REVIEWERS

The Journal Editorial Team would like to thank the reviewers for their time and effort. The comments that we received were very constructive and detailed, and help us to continue to produce a consistently top-quality journal. Your participation is very important in the success of providing a distinguished outlet for original valuable articles. Again I would like to thank you all for your assistance in the review process.

Sandy Furterer, Ph.D., MBA
Editor-in-Chief

TABLE OF CONTENTS

Tracking Projectile Motion with a Wireless Sensor

Eunice E. Yang[1]
[1]*University of Pittsburgh at Johnstown*
eyang@pitt.edu

Abstract

A wireless sensor containing an altimeter and accelerometer is used to track the projectile motion of a foam soccer ball. Theory of projectile motions with and without drag are presented. Theoretical and experimental comparisons are made with regards to altitude and acceleration of the soccer ball. The exercises presented in this paper are appropriate for undergraduate students enrolled in introductory physics or engineering mechanics dynamics courses. The exercises were completed by students to help them achieve better understanding of physics concepts that are troublesome to many.

1. Introduction

Tracking the flight of various athletic balls have been previously investigated using electronic sensors[1] and video analysis[2]. In the past, a background in electronics and programming was needed to use these electronic sensors. In addition, tracking flights using videos required expensive high speed cameras. More recently, the availability of simple and cost effective sensors has made it possible for students without much background in electronics or programming to monitor flight. This paper discusses how a sensor can be used to help students gain a better understanding of projectile motion.

In introductory physics and dynamics courses, air resistance is neglected so that projectiles of objects such as a baseball or even a motorcycle are treated as particles. This approach has its benefits and drawbacks in that students can find solutions analytically but the pedagogical opportunity to help them realize how significantly higher theoretical solutions can be compared to experimental results is lost.

In this paper, a cost-effective stand-alone wireless sensor is used to capture altitude and acceleration data of a foam soccer ball in projectile motion. The data from the sensor are transmitted wirelessly to an app on a smartphone or iPad® which then can be exported as a file for analysis in a spreadsheet software. This hands-on exercise helps students to avoid the misconception that all rigid bodies in projectile motion have acceleration equivalent to the gravitational constant. In addition, they gain a better understanding of how experimental values compare to theoretical predictions.

2. Experimental Method

A foam soccer ball was thrown and its altitude and acceleration were tracked using a wireless Bluetooth® sensor (Figure 1) [3]. The sensor contained a Bosch BMP 1180 digital pressure sensor to measure altitude as a function of atmospheric pressure and a 3-axis STMicroelectronics LIS331HH accelerometer. It measured 49 mm x 18 mm x 14.5 mm in size and weighed 10.5 grams. Altitude data was acquired every 0.05 seconds with a resolution of ±0.25 m. Acceleration was acquired every 0.005 seconds with a resolution of ±0.1 m*g*. A soccer ball measuring 19 cm in diameter and weighing 0.235 kg was cut in half and the sensor was placed in the center. Four straight channels 90° apart were made so that the pressure sensor could detect the change in atmospheric pressure (Figure 2).

Figure 1. The wireless sensor shown next to a U.S. quarter to show relative size.

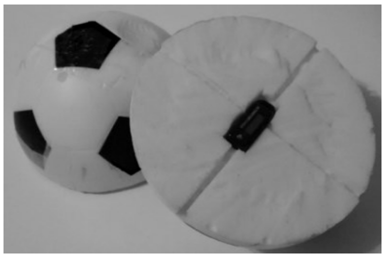


Figure 2. Foam soccer ball with the wireless Bluetooth® sensor. The sensor was capable of acquiring altitude and acceleration about 3-axes.

An attractive feature of this particular sensor was that it did not require any filtering, smoothing, programming, or mathematical manipulation of data by the students. In addition, using the manufacturer's app that was installed on a smart device allowed students to view a graphical plot of altitude and acceleration as a function of time immediately after each flight (Figure 3). Data was easily accessible since the app provided CSV formatted data files to be uploaded to an email account or to cloud storage for later analysis.

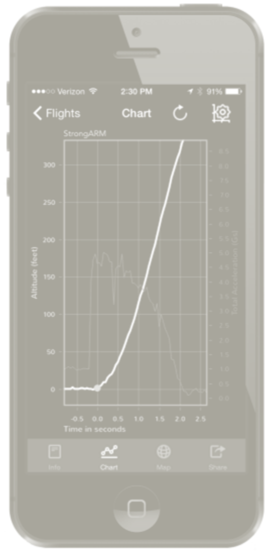


Figure *3*. Image of the altitude and total acceleration that could be viewed after each flight using the manufacturer's app installed on a smartphone.

An iPad® recording at a rate of 30 frames per second was used to video record the initial portion of each throw. The cost-free video analysis and modeling software Tracker[4] popular among many general physics instructors was utilized to determine the initial velocity and angle of throw.

1. Projectile Motion Theory

Kinematic equations governing projectile motions are taught in introductory physics and dynamics courses. The objects, regardless of size, are idealized as a particle and simplify the analysis since air resistance can be neglected.

This idealization allows the student to assume acceleration in the horizontal and vertical directions to be zero and -9.81 m/s² (or -32.2 ft/s²), respectively. Kinematic equations for projectile motion with negligible drag are typically presented in dynamics[5] or general physics[6] textbooks as

$$x = x_o + v_o t, \quad (1)$$

$$v_y = (v_o)_y - g, \quad (2)$$

$$y = y_o + (v_o)_y\, t - \tfrac{1}{2} g t^2, \quad (3)$$

and

$$v_y^2 = (v_o)_y^2 - 2g(y - y_o) \quad (4)$$

where x_o and y_o are the initial position coordinates, x and y are the positions at time t, $(v_o)_x$ and $(v_o)_y$ are the initial velocity components. A closed form solution to these equations is easily obtainable by students.

The free-body diagram of a rigid-body of mass m that includes the effects of drag force F_D is shown in Figure 4.

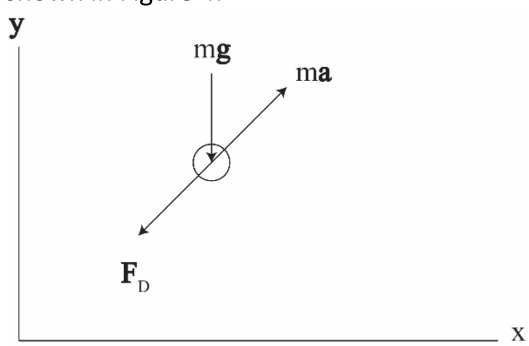


Figure 4. Free-body diagram of a rigid body in projectile motion.

Applying Newton's second law in the x- and y- directions yield

$$ma_x-(F_D)_x=0 \quad (5)$$

and

$$-mg-(F_D)_y=ma_y, \quad (6)$$

respectively. For velocities where the Reynolds number is between 10^4 and 10^5, the drag force can be approximated as[7]

$$F_D=\frac{1}{2}\rho C_D A v^2 \quad (7)$$

where ρ is the density of air, C_D is the drag coefficient of the body, A is the frontal area, and v is the speed of the body. Substitution of Eq. (7) into Eqs. (5) and (6) and then solving for a_x and a_y yield

$$a_x=\frac{\frac{1}{2}\rho C_D A v_x^2}{m} \quad (8)$$

and

$$a_y=\frac{-mg-\frac{1}{2}\rho C_D A v_y^2}{m}, \quad (9)$$

respectively. The velocity components are given by

$$v_x=dx/dt \quad (10)$$

and

$$v_y=dy/dt. \quad (11)$$

Eqs. (8-11) are the kinematic equations that describe the motion of the rigid body. However, no closed form solution exists and must be solved numerically. Euler's method of forward time differentiation[8,9] is one method of choice and can simply be implemented using computational software such as Microsoft Excel[10], MathCAD, or MATLAB[11].

2. Results and Discussion

The foam soccer ball was thrown twice at "slow" and "fast" speeds of 10.9 m/s and 19.0 m/s, respectively. Initial conditions of these two throws were obtained by video analysis using Tracker and are shown in Table 1.

Table 1. Initial conditions of two trials of a soccer ball set in projectile motion at "slow" and "fast" speeds.

	$(v_x)_1$ m/s	$(v_y)_1$ m/s	x_1 m	y_1 m	θ_1 deg
"Slow"	3.73	10.24	0	0.9	70.0
"Fast"	13.52	13.87	0	1.2	44.7

In Figure 5, altitude (closed circle "●") as a function of the time of both throws are compared to theoretical predictions of altitude with (thin solid line) and without drag (thick solid line) as a function of time. When students see these comparisons they begin to realize the significant error of ignoring the effects of drag.

Table 2 compares the theoretical predictions of maximum altitude of both throws to experimental data. Both theories over predict experimental data. For the fast soccer ball, maximum altitude errors for predictions with and without drag are +47% and +88%, respectively. It should be noted that the asymmetrical placement of the sensor in the ball may have affected flight performance and contributed to the resulting lower flight altitudes.

Upon further analysis, students begin to wonder why theoretical predictions with drag is still +47% and +35% greater than experimental results for the slow and fast throws, respectively. For a possible explanation, they are asked to consider the effects of the rotation of the ball on altitude. The rotational characteristics of the soccer ball are captured by the 3-axis accelerometer (open circle "**o**") and are shown in Figure 5. In introductory physics and dynamics, students are taught to assume that the total acceleration for a rigid body assumed to be a particle is 1*g*.

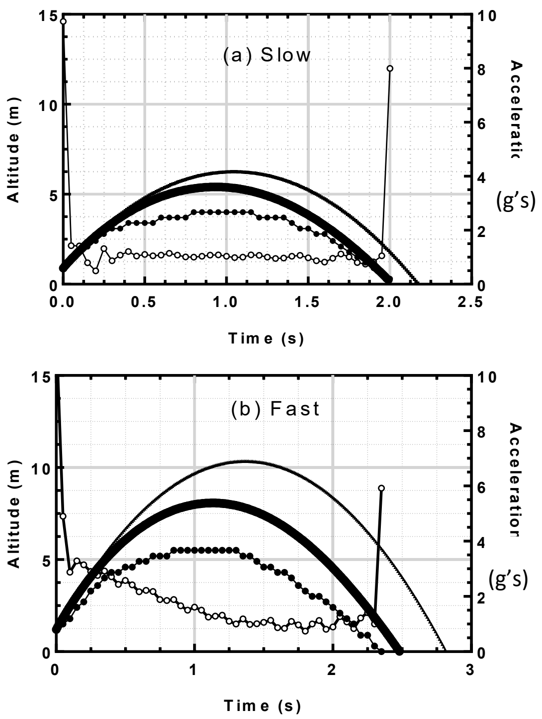


Figure 5. Comparison of theoretical values to experimental data of altitude and total acceleration as a function of time for (a) slow and (b) fast throwing conditions. (Theory no drag "—"; Theory with drag "▬ " ; Experimental altitude "●"; Experimental acceleration "o")

Table 2. Comparison of experimental altitude and distance travelled to theoretical values.

	Fast		Slow	
	y_{max}	Error	y_{max}	Error
Experimental	5.5	-	4.0	-
Theory (drag)	8.1	+47%	5.4	+35%
Theory (no drag)	10.0	+88%	6.3	+56%

Figure 5 shows that the slow ball has an average total acceleration of approximately 1*g* throughout its flight. However, the fast ball exhibits a total acceleration that is much higher at 3.3*g's* at the onset of the throw. After it has reached its maximum height the acceleration maintains 1*g* acceleration. The higher total acceleration prior to reaching maximum height is due to the spin of the soccer ball which also attributes to the larger error in maximum altitude y_{max} shown in Table 2.

5. Conclusion

Use of wireless sensors to measure altitude and acceleration of a rigid body in projectile motion provide pedagogical opportunities to help students avoid misconceptions and better understanding of basic physics concepts. The wireless sensor was placed in the center of a foam soccer ball and set into projectile motion. A soccer ball with an initial velocity of 19.0 m/s, had theoretical predictions of maximum altitude that was 88% (numerical prediction without drag) greater than experimental results. A soccer ball that was thrown at a slower speed ball of 10.9 m/s had a slightly lower error of +56% (numerical prediction without drag). The reason for this difference was the spinning that the faster ball experienced. This was evident to the students as the total acceleration of the fast ball was greater than 1 *g* while the slower ball's total acceleration remained equivalent to the gravitational constant *g*.

A more advanced use of cost effective sensors can be extended to undergraduates studying missile guidance system[12] perhaps in an aerospace engineering curriculum. Instead of the 3- degrees of freedom accelerometer used in this paper a 9- degrees of freedom inertial measurement unit (IMU) sensor could be used.

The IMU contains a 3-axis accelerometer, 3-axis gyroscope and 3-axis magnetometer to monitor vibration, angle of rotation, and Earth's magnetic field, respectively. Other advanced applications using IMUs include design of quadrotors[13] and robotics[14] to name a few.

6. References

[1]C. J. Nowak, "Flight Data Recorder for the American Football," State University of New York at Buffalo (2003).

[2]G. K. Hung, B. Johnson, and A. Coppa, "Aerodynamics and biomechanics of the free throw", in *Biomedical Engineering Principles in Sports* (Springer, 2004), pp. 367-390.

[3]https://www.jollylogic.com, AltimeterThree, (2016).

[4]D. Brown, Tracker, <http://physlets.org/tracker/> (2015).

[5]R. C. Hibbeler, *Engineering mechanics*, (Pearson, Singapore;Hoboken, NJ;, 2016).

[6]S. T. Thornton and A. F. Rex, *Modern physics for scientists and engineers*, (Cengage Learning, Boston, MA, 2013).

[7]S. F. Hoerner, *Fluid-dynamic Drag: Practical Information on Aerodynamic Drag and Hydrodynamic Resistence*, (Sighard F. Hoerner, 1965).

[8]G. A. Finney, "Analysis of a water-propelled rocket: a problem in honors physics," American Journal of Physics [H.W.Wilson - AST]. **68** (3), 223 (2000).

[9]D. Kagan, "Soda-Bottle Water Rockets," Physics Teacher. **33** (3), 150 (1995).

[10]J. Benacka, "Simulating projectile motion in the air with spreadsheets," Spreadsheets in Education (eJSiE). **3** (2), 3 (2009).

[11]A. C. Dhanemozhi, "Projectile motion using MATLAB," International Journal of Science, Technology & Management. **04** (Special Issue 1), 1103-1107 (2015).

[12]R. R. Minor and D. W. Rowe, "Utilization of GPS/MEMS-IMU for measurement of dynamics for range testing of missiles and rockets," in *IEEE 1998 Position Location and Navigation Symposium (Cat. No.98CH36153)*. 1998). 602-607.

[13]S. Gupte, M. Paul Infant Teenu, and J. M. Conrad, "A survey of quadrotor Unmanned Aerial Vehicles," in *2012 Proceedings of IEEE Southeastcon*. 2012). 1-6.

[14]G. Santaera, et al., "Low-cost, fast and accurate reconstruction of robotic and human postures via IMU measurements," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. 2015). 2728-2735.

Off-line Optimization for Aerial Spray to Fields with Non-uniform Needs

Y. Zhang

Department of Agriculture Mechanization and Automation

Heilongjiang Bayi Agriculture University
P. R. China

Wei Zhan
Department of Engineering Technology, Texas A&M University, College Station, TX 77843.
Corresponding email: wei.zhan@tamu.edu

Abstract

Aerially applied spray of chemicals to crops in fields is an effective way of applying chemicals to the fields for different needs. The chemicals can be fertilizers and treatment for weeds. One of the main challenges in aerial spray is to minimize the downwind drift of the chemicals, which can cause pollutions and harm to crops in nearby fields. Simulation tools such as AGDISP are widely used for prediction of downwind spray under various conditions. In this paper, a new off-line optimization approach is proposed to minimize the downwind drift of chemicals for areas with non-uniform needs. Aerial photos for the target area will be taken. The non-uniform needs will then be analyzed using image process methods. Off-line simulation will be run to optimize the downwind drift. Based on previous research results, wind speed, release height, and droplet size will be the main factors that will be studied. The amount of spray will be controlled by adjusting the mixture of water and chemicals along the flight trajectory. The swath width, flight speed, and other factors will be studied during off-line simulation to evaluate their impact on the downwind drift. Optimal values for these variables will be recommended. This new scheme of aerial spray planning requires research efforts in image processing, on-board control system design, AGDISP simulation, and off-line optimization.

1. Introduction

There are many ways of applying fertilizers and pesticides to crops in fields for different needs. Aerially applied spray, illustrated in Fig. 1, is one of the effective methods. For crops taller than a certain height, aerial spray may be the only option.

Ideally, one would like to have all the sprayed chemicals to be applied to the crops and none should go out of the desired area. However, in practice, there is always a small amount of the sprayed chemicals that would miss the target. This is particularly the case when there is wind. Downwind drift of the chemicals can cause pollutions and harm to crops in nearby fields (Yates *et al.*, 1967; Cooper and Alley, 1994; Bird *et al.*, 1996; Seinfeld and Pandis, 1998). Some of such pollution from the pesticide spray can cause cancer.

Figure 1. Aerial spray of fertilizers or pesticides

Off-target movement of aerially applied chemicals is a well-researched topic in the last five decades (Pasquill, 1961, Yates *et al*., 1966; Yates *et al*., 1967; Pasquill and Smith, 1983; Bird, 1995; Ganzlemeier *et al*., 1995; Arvidsson, 1997; Smith *et al*., 2000; Hewitt *et al*., 2001a; Maber *et al*., 2001; Fritz *et al*., 2008b; Huang *et al*, 2010). The focus of the research in this area is on minimizing the downwind drift of the chemicals (Turner, 1994; Teske *et al*., 2002; Teske *et al*., 2003). There are various metrics that can be used to measure the success of aerial sprays, these include the spray efficiency, downwind deposits at specific distances away from the target area (EPA, 1998; EPA, 2000; ASABE, 2004).

Downwind drift of aerial spray is affected by many factors including wind speed and direction (Pasquill, 1961), temperature (Fritz *et al.*, 2008), humidity, release height, nozzle configuration, type of airplane, canopy structure (Praat *et al*., 2000), droplet size (Hewitt *et al*., 1996), tank mix (Hewitt *et al.*, 2001b), plant structure (Praat *et al*., 2000), air speed (Fritz *et al*., 2008a), and other factors (Hewitt *et al*., 2001c; Thistle, 2000). The objective is to find a way to select the values for factors that the applicator can change such as release height and droplet size to minimize the impact of the factors that cannot be changed, such as temperature and humidity. In addition to optimizing the values for certain factors, other methods such as electrostatic spray were also studied (Carlton *et al*., 1995; Shu, 2012).

This paper discusses the optimization of aerial spray to fields with non-uniform needs. The basic approach is using off-line simulation to optimize the spray. This is a continuation of research work conducted by Huang *et al*. (Huang *et al.,* 2010; Huang *et al*., 2012).

The remainder of the paper is organized as follows. Section 2 introduces the off-line simulation technique using software AGDISP. A new optimization method is presented in Section 3. Conclusions and future work are discussed in Section 4.

2. Off-line optimization using AGDISP

Research work in aerial sprays can be divided into two types: field test (Pasquill, 1961; Yates *et al*., 1966; Bird, 1995; SDTF, 1997; Praat *et al*., 2000; Fritz, 2006; Fritz *et al*., 2008b) and simulation (Akesson *et al*., 1981; Walklate, 1987; Bilanin *et al*., 1989; Turner, 1994; Zhu *et al*., 1995; Kaul *et al*., 1996; Potter *et al*., 2000; Hewitt *et al*., 2002; Teske *et al*., 2002; Fritz *et al*., 2008a). Statistical analysis of field test data can reveal the impact of various factors. Modeled-based simulation can provide a cost effective way to predict results of aerial spray. Field test is time consuming, costly, but is necessary for validating other prediction methods. The simulation approach provides more flexibility in terms of setting values for various factors, cost, and execution time; however, the results depend very much on the accuracy of the model.

AGDISP (AGricultural DISPersion) (Bilanin *et al*., 1989) is a software package widely used to study the impact of various factors on spray material movement. AGDISP has been validated with test data by many researchers over the last 30 years (Teske and Barry, 1993; Bird *et al*., 2002; Huang *et al.,* 2010; Huang *et al*., 2012). The graphical user interface (GUI) is shown in Fig. 2. All simulation parameters can be changed, some accepting numerical values and others with pulldown menus. There are building limits for some of the parameters. For example, the highest temperature is 125 F. An error message will pop up if a value outside the range is selected.

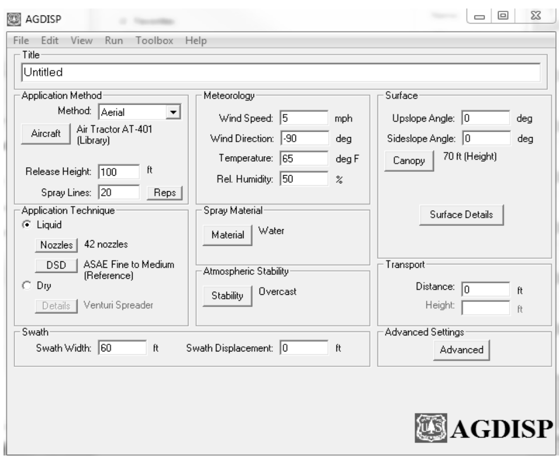


Figure 2. AGDISP simulation GUI

Using AGDISP simulation, one can predict downwind deposition under specific test conditions. Improvement action can be recommended based on the simulation result and additional data analysis. Fig. 3 illustrates the comparison of results with default swath offset value (solid line) and optimized swath offset value (dotted line). The optimal swath offset value is calculated in MATLAB using the data exported from AGDISP.

Wingman (ADPACO, 2011), a commercial product developed by ADAPCO, uses AGDISP to optimize the flight trajectory in situ in cockpit. The drawback for such in situ optimization includes high cost equipment and insufficient time for optimization. While the detail of Wingman algorithm is not available to public, the time it takes to execute the algorithm makes the result not absolutely optimal, instead some sub-optimal result will be achieved.

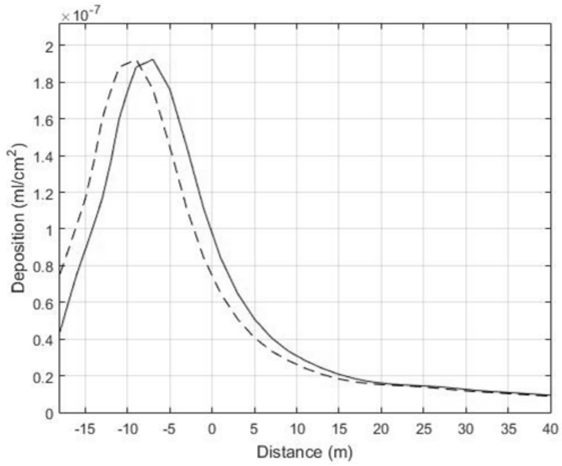


Figure 3. A typical use of AGDISP

Huang *et al.* (2012) proposed an alternative to Wingman: an offline optimization process that does not require in situ control algorithm. The basic idea for this off-line optimization method is that of all the factors, only a few have a major impact on the result, therefore one can focus on these main factors to find a sub-optimal solution. With the off-line optimization, charts and tables will be generated beforehand for applicators instead of using the expensive in-situ cockpit optimizer.

Fig. 4 illustrates the curve-fitting function based on off-line optimization using AGDISP and MATLAB. For certain droplet size (very coarse to extremely coarse) and release height (1.9 m), the sub-optimal swath offset can be calculated as a function of wind speed.

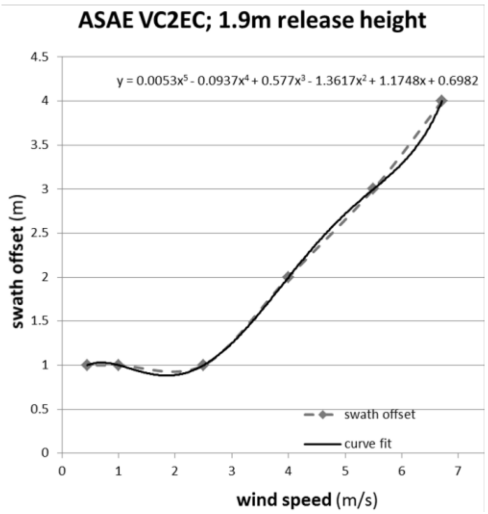


Figure 4. Sub-optimal swath offset as a function of wind speed

In addition to its low-cost advantage over Wingman, the off-line optimization method also provides applicators a tool to make a decision on whether to spray under specific weather conditions.

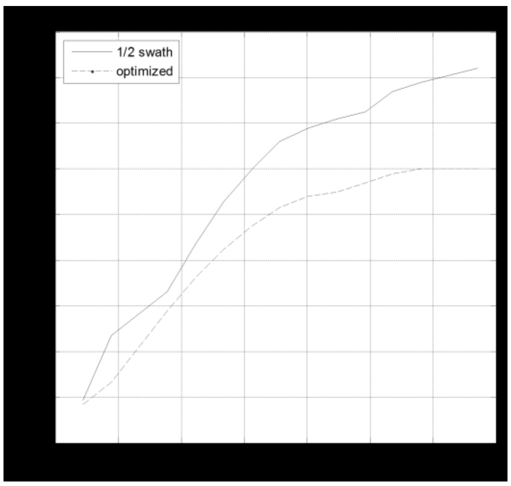


Figure 5. Use of off-line optimization result

Fig. 5 illustrates the downwind deposition between 30.48 m and 45.72 m as a function of wind speed. If the applicator wants to have the deposition level below 1.2 ml/cm^2 for example, then he should only spray when the wind speed is below 3.5 m/s. A quick check on the weather forcast would enable the applicator to make a decision on whether to spray or not. This is a significant advantage over Wingman, which only tries to get the best result for a given condition with all the sensors taking measurement during the flight. The dotted line in Fig. 5 represents the optimized result and the solid line represents the result if the default value of ½ swath is used.

3. New method for downwind drift optimization

Inspired by the work of Huang *et al* (2012) an off-line optimization to improve the spray efficiency, this paper proposes to use off-line optimization methods on non-uniform spray areas. The current spray practice and AGDISP simulation assume that the spray need for a given area is uniform. A non-uniform spray area is illustrated in Fig. 6, where there are more weeds in the highlighted areas. It is common that certain parts of the area need higher levels of spray than the rest.

Figure 6. Non-uniform spray area

Apparently, uniform spray would cause either certain areas not getting enough spray or other areas getting too much spray as well as increased downwind deposition of chemicals. In this paper, the following two steps are proposed to improve the spray efficiency.

- Using AGDISP first followed by off-line optimization in MATLAB;
- Using an automatic control system with GPS and off-line optimization result as inputs to change the spray amount as needed during flight.

In the first step, the aerial photo of the non-uniform area is taken and analyzed. To illustrate the idea, a simple case of two level spray needed is considered as illustrated in Fig. 7, where the arrow is added to indicate the horizontal axis that will be used later for plotting the downwind deposition. The darker sub area needs twice as much spray than the rest of the area.

Spray parameters are selected in AGDISP so that the average deposition in the target area meets the requirement for the darker area in Fig.7.

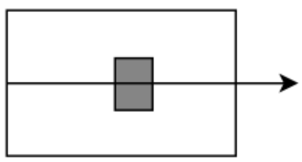


Figure 7. Characterization of spray needs for different parts of the target area

This may take a few trials. Since this step is carried out off-line, simulation time is not an issue. The simulation result is then exported to MATLAB for optimization. Fig. 8 shows the deposition curve for the area specified in Fig. 7 based on the AGDISP simulation when the spray does not take the non-uniform need into consideration.

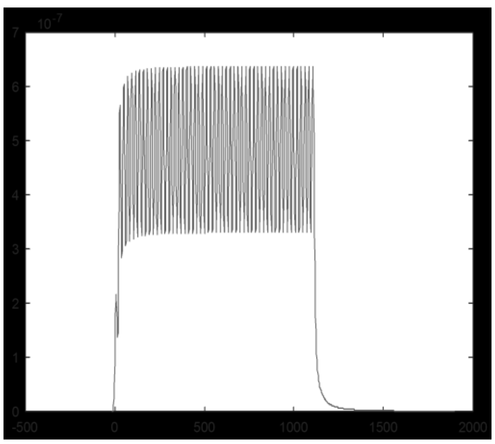


Figure 8. AGDISP simulation data

By combining AGDISP and MATLAB, the features of GDISP are enhanced for non-uniform spray. The resulting deposition is plotted in Fig. 9. The horizontal axis in Fig. 9 represents the distance from the left edge of the whole area as shown in Fig. 7. As one can see in the figure, in the middle area the deposition is about twice as high as the rest of the distance.

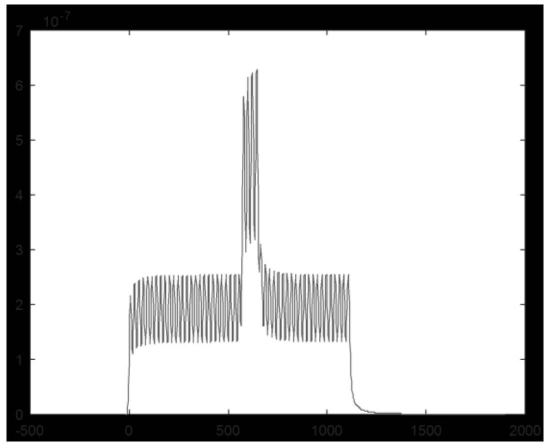


Figure 9. Deposition with non-uniform spray

With the off-line optimization completed, the next step is to control the spray amount according to the GPS information. In situ flow control is currently not available during aerial spray unless some advanced control systems such as Wingman (ADAPCO, 2011) are used. One of the methods for controlling the spray amount is illustrated in Fig. 10. Two containers are connected to the spray nozzle through a switchable control valve. Each container has chemical solutions with different concentration levels.

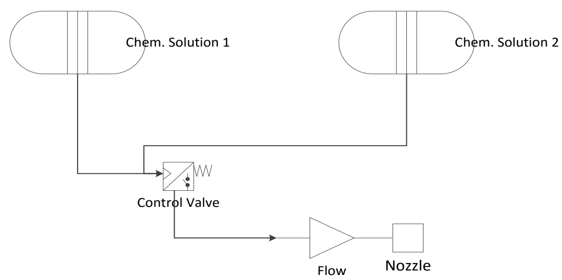


Figure 10. Control of concentration level

This control system together with a GPS input would allow one to switch between the two concentration levels during the flight as needed. As illustrated in Fig. 11, when the flight is at point A the control system would switch from low concentration level to high concentration level and switch back to low level at point B. This will continue until point C and D are passed. After that, low level will be used. Conceptual wise, this spray method should work better than the uniform spray method that every applicant is using.

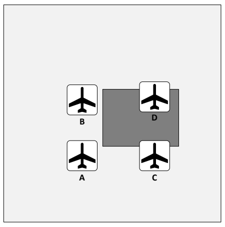


Figure 11. GPS controlled concentration level switching

Of course, the real situation can be a lot more complicated than what is illustrated in Fig. 7. One may need more than two different levels of concentrations. The off-line simulation may become too complicated. It may come down to a trade-off between spray efficiency and computational effort.

The setup in Fig. 10 may need to be modified when many different concentration levels are needed. One option is to have two valves that control the flowrates of the water and the chemicals. If practical issues like this can be successfully solved, then the potential improvement of the new spray method over the existing ones can be great.

4. Conclusions and future work

A new off-line optimization approach is proposed to reduce the downwind drift of chemicals for areas with non-uniform needs.

Aerial photos for the target area will be taken. The non-uniform needs will then be analyzed using image process methods. The amount of spray will be controlled by adjusting the mixture of water and chemicals or switch between different levels of concentrations with pre-mixed chemical solutions along the flight trajectory with feedback from GPS. A combination use of AGDISP and MATLAB can enhance the AGDISP simulation capability so that it can be used for non-uniform sprays. Although adjustable flow control is needed by the method proposed in this paper, one has the option of using or not using other in situ spray control system such as Wingman. The focus of this paper is the off-line optimization method.

Future work includes generalization of results in (Huang *et al.*, 2012) so that main factors such as wind speed, swath offset, release height, and droplet size will be studied. Optimal values for these variables will be recommended based on off-line simulation. The implementation of control systems is the next subject of research in order for the new spray method to work as proposed in this paper. The new spray method would then be validated in field testing.

References

ADAPCO. 2011. WINGMAN User's Manual. http://www.myadapco.com/res/pdf/manuals/Wingman_Manual.pdf

Akesson, N.B., Yates, W.E., Smith, N., and Cowden, R.E.. 1981. Rationalization of pesticide drift-loss accountancy by regression models. ABAE Paper No. 81-1006, ASAE, St. Joseph, MI.

Arvidsson, T. 1997. Spray drift as influenced by meteorological and technical factors. A methodological study. Swedish University of Agricultural Sciences, Acta Universitatis Agriculturae Sueciae, Agraria 71.

ASABE. 2004. S561.1, Procedures for measuring drift deposits from ground, orchard, and aerial sprayers. ASABE, St. Joseph, MI.

Bilanin, A.J., Teske, M.E., Barry, J.W., and Ekblad, R.B. 1989. AGDISP: The aircraft spray dispersion model, code development and experimental validation. " Transactions of the ASAE 32: 327-334.

Bird, S.L. 1995. A compilation of aerial spray drift field study data for low-flight agricultural application of pesticides. In Environmental Fate of Agrochemicals: A Modern Perspective. ed. M.L. Leng, E.M.K Loevey, and P.L. Zubkoff, Lewis Publishers, Chelsea, MI.

Bird, S.L., Esterly, D.M., and Perry, S.G. 1996. Atmospheric pollutants and trace gases. J. Environ. Quality 25(5): 1095-1104.

Bird, S.L., Perry, S.G., Ray, S.L., and Teske, M.E., 2002. Evaluation of the AgDISP aerial spray algorithms in the AgDRIFT model. Environmental Toxicology and Chemistry 21(3): 672-681.

Bouse, L.F. 1994. Effect of nozzle type and operation on spray droplet size. Transactions of the ASAE 37(5):1389-1400.

Carlton, J. B., Bouse, L. F., & Kirk, I. W. (1995). Electrostatic charging of aerial spray over cotton. Trans. of the American Soc. of Agric. Eng., 38(6), 1641-1645.

Cooper, C.D. and Alley, F.C. 1994. Air Pollution Control: A Design Approach. 2nd Edition. Waveland Press, Inc., Prospect Heights, IL.

Environmental Protection Agency. 1998. Spray drift test guidelines – Spray drift field deposition. OPPTS 840.1200, EPA 712-C-98-112, USEPA Office of Prevention, Pesticides and Toxic Substances (OPPTS), Washington DC.

Environmental Protection Agency, 2000. Meteorological monitoring guidance for regulatory modeling applications. EPA-454/R-99-005, USEPA Office of Air Quality Planning and Standards, Research Triangle Park, NC.

Fritz, B.K. 2006. Meteorological effects on deposition and drift of aerially applied sprays. Transactions of the ASABE 49(5): 1295-1301.

Fritz, B.K., Bagley, B., Hoffmann, W.C., and Lan, Y. 2008a. Spray spectrum modifications through changes in airspeed to minimize drift. 41th Annual National Agricultural Aviation Association Convention, paper No. AA08-002, Las Vegas, NV,

Fritz, B.K., Hoffmann, W.C., Lan, Y., Thomson, S.J., and Huang, Y. 2008b. Low-level atmospheric temperature inversions: characteristics and impacts on aerial applications. Agr. Engng. Intl. Vol. X. Manuscript PM 08 001.

Ganzlemeier, H., Rautmann, D., Spangenberg, R., Streloke, M., Herrmann, M., Wenzelburger, H., and Walte, H. 1995. Studies on the Spray Drift of Plant Protection Products. Blackwell Wissenschafts-Verlag GmbH Berlin/ Wien, Berlin.

Hewitt, A.J., Valcore, D.L., and Bryant, J.E. 1996. Spray drift task force atomization droplet size spectra measurements. Proceedings ILASS-Americas 96, San Francisco, CA.

Hewitt, A.J., Johnson, D., Fish, J.D., Hermansky, C.G., and Valcore, D.L. 2001a. The development of the spray drift task force database on pesticide movements for aerial agricultural spray applications. Environmental Toxicology and Chemistry 21 (3): 648-658.

Hewitt, A.J., Miller, P.C.H., Dexter, R.W., and Bagley, W.E. 2001b. The influence of tank mix adjuvants on the formation, characteristics and drift potential of agricultural sprays. International Symposium on Adjuvants for Agrochemicals, Amsterdam, Netherlands.

Hewitt, A.J., Valcore, D.L., and Barry, T. 2001c. Analyses of equipment, weather and other factors affecting drift from applications of sprays by ground platforms. Pesticide Formulations and Applications Systems 20-ASTM STP 1400: 44-58.

Hewitt, A.J., Maber, J., and Praat, J.P. 2002. Drift management using modeling and GIS systems. Proceedings of the World Congress of Computers in Agriculture and Natural Resources, 290-296.

Hoffmann, W.C. and Salyani, M. 1996. Spray deposition on citrus canopies under different meteorological conditions. Transactions of the ASAE 39(1): 17-22.

Hoffmann, W.C. and Tom, H.H. 2000. Effects of lowering spray boom in flight on swath width and drift. Appl. Engr. Agric. 16(3): 217-220.

Huang, Y., Zhan, W., Fritz, B.K., Thomson, S.J., and Fang, A. 2010. Analysis of impact of various factors on downwind deposition using a simulation method. Journal of ASTM International 7(6): 1-11.

Huang, Y., Zhan, W., Fritz, B., and Thomson, S., 2012. "Optimal selection of controllable variables to minimize downwind drift from aerially applied sprays," Applied Engineering in Agriculture, 28(3), pp. 307-314.

Kaul, P. von, Gebauer, S., Neukampf, R., and Ganzelmeier, H. 1996. Modeling of direct drift of plant protection products – field sprayers. Nachrichtenbl. Deut. Pflanzenschutzd 48(2): 21-31.

Maber, J., Dewar, P., Praat, J.P., and Hewitt, A.J. 2001. Real time spray drift prediction. Proc. IT Symposium, New Zealand.

Miller, P.C.H., Lane, A.G., Walklate, P.J., and Richardson, G.M. 2000a. The effect of plant structure on the drift of pesticides at field boundaries. Aspects of Applied Biology 57: 75-82.

Pasquill, F. 1961. The estimation of the dispersion of windborne material. Meteorological Mag. 90: 33-49.

Pasquill, F. and Smith, F.B. 1983. Atmospheric Diffusion: A Study of the Dispersion of Windborne Material from Industrial and Other Sources. 3rd Ed., Ellis Horwood Limited, Chichester, West Sussex, England.

Potter, W.D., Bi., W., Twardus, D., Thistle, H.W., Ghent, J., Twery, M., and Teske, M.E. 2000. A genetic algorithm for aerial spray application optimization. ASAE Paper No. 001053, ASAE, St Joseph, MI.

Praat, J.P., Maber, J., and Manktelow, D.W.L. 2000. The effect of canopy development and sprayer position on spray drift from a pipfruit orchard. Proc NZ Plant Protection, 241-247.

Seinfeld, J.H. and Pandis, S.N. 1998. Atmospheric Chemistry and Physics – From Air Pollution to Climate Change. John Wiley & Sons, Inc. Hoboken, NJ.

Shu, C. 2012. Deposition effects of electrostatic and non-electrostatic aerial spray. Forest Sci., 48(4), 75- 79.

Smith, D.B., Bode, L.E., and Gerard, P.D. 2000. Predicting ground boom spray drift. Transactions of the ASAE 43(3): 547-53.

Spray Drift Task Force. 1997. A Summary of Aerial Application Studies. http://www.agdrift.com/PDF_FILES/Aerial.pdf

Teske, M.E., Bird, S.L., Esterly, D.M., Curbishley, T.M., Ray, S.L., and Perry, S.G. 2002. AgDRIFT®: A model for estimating near-field spray drift from aerial applications. Environmental Toxicology and Chemistry 21(3): 659-671.

Teske, M.E., Thistle, H.W., and Ice, G.G. 2003. Technical advances in modeling aerially applied sprays. Transactions of the ASABE 46(4): 985-996.

Thistle, H.W. 2000. The role of stability in fine pesticide droplet dispersion in the atmosphere: a review of physical concepts. Transactions of the ASAE 43(6): 1409-1413.

Turner, D.B. 1994. Workbook of Atmospheric Dispersion Estimates: An Introduction to Dispersion Modeling. 2nd ed., CRC Press, Boca Raton, FL.

Walklate, P.J. 1987. A Random-Walk model for dispersion of heavy particles in turbulent air flow. Boundary-Layer Meteorology 39: 175-190.

Yates, W.E., Akesson, N.B., and Coutts, H.H. 1966. Evaluation of drift residues from aerial applications. Transactions of the ASAE 9(3): 389-393.

Yates, W.E., Akesson, N.B., and Coutts, H.H. 1967. Drift hazards related to ultra-low-volume and diluted sprays applied by agricultural aircraft. Transactions of the ASAE 10(5): 628-632.

Zhu, H., Reichard, D.L., Fox, R.D., Ozkan, H.E., and Brazee, R.D. 1995. DRIFTSIM, A program to estimate drift distances of spray droplets. Applied Engineering in Agriculture 11(3): 365-369.

Is the End of too Big to Fail the Start of Something Worse?

Jarrett Kotrozo[1]
Edward H. Hernandez[2]
[1]*California State University, Stanislaus*
[2]*California State University, Stanislaus*
jkotrozo@csustan.edu

Abstract

Passage of the Dodd-Frank Wall Street Reform and Consumer Protection Act in 2010 effectively put an end to the doctrine of "too big to fail" (TBTF). In doing so, the legislation increases the costs of doing business for financial services firms, particularly bank holding companies. The legislation also has an impact on corporate governance mechanisms, specifically the board of directors, stakeholder theory and agency theory. In addition to its objective of reducing moral hazard, unintended consequences of the legislation may adversely affect shareholder protection and dampen economic growth in the United States.

1. Introduction

"I … emphasized to Ken [Lewis, Bank of America CEO] that the government would not let any systematically important institution fail; that exercising the MAC [Material Adverse Change Clause] would show a colossal lack of judgment by BofA; that such an action would jeopardize [BofA], Merrill Lynch, and the entire financial system; and that under such circumstances, the Fed, as BofA's regulator, could take extreme measures, including the removal of management and the board" (Paulson, 2010 p. 429).

The exact intent of these words, uttered by U.S. Treasury Secretary Henry "Hank" Paulson, to Bank of America CEO, Ken Lewis, during the financial crisis of 2008, may never be known, but the intent was clear. Secretary Paulson, in his attempt to rescue the U.S. financial system from potential collapse, suggested the possibility of using his regulatory power as U.S. Treasury Secretary to remove a CEO and his board for expressing reservations about consummating a merger between Bank of America and Merrill Lynch, in light of growing losses at Merrill Lynch.

After reporting a $5.1 billion loss in the third quarter of 2008, a November 12, 2008 internal Merrill forecast predicted an additional loss of $5 billion in the fourth quarter, just days before the merger was to be finalized on January 1, 2009 (statement of Timothy J. Mayopolous before the U.S. House of Representatives Committee on Oversight and Government Reform, 2009). These losses soared to an expected $22 billion according to Bank of America staff on December 19, 2008 including an additional $4 billion loss in just 48 hours. This escalation led CEO Lewis and his board to question whether this merger was in the best interest of the organization, which is precisely the fiduciary role boards of directors are expected to play, under the law. However, the importance of these "systematically important" institutions may have trumped fiduciary duty due to the doctrine of "too big to fail." (Interested readers may wish to reference Harris and Kotrozo's, "Bank of America's acquisition of Merrill Lynch: A Multi-Theoretical Analysis" 2010 *Journal of Management and Engineering Integration* 3(1) pp. 21-32 for a more in-depth analysis.)

While a topic of discussion in academic literature for some time, too big to fail (TBTF) entered the national and international lexicon as a result of the 2008 financial crisis. With the inherent, and now more readily-perceived unfairness of the moral hazard which goes hand-in-hand with TBTF, this discussion left the realm of academe and made its way into international regulatory agencies such as the Basel Committee on Banking and Supervision as part of the Basel III accords, and with the United States Congress. Specifically, in July 2010 Congress passed, and President Obama signed into law, the Dodd-Frank Wall Street Reform and Consumer Protection Act to, "[P]romote the financial stability of the United States by improving accountability and transparency in the financial system, to end "too big to fail," to protect the American taxpayer by ending bailouts, to protect consumers from abusive financial services practices, and for other purposes" (Dodd-Frank Act, 2010 p. 1).

It is the end of the TBTF doctrine which is the focus of this paper. Although the desire to curb the moral hazard inherent in TBTF is a noble endeavor, the cure, as laid out in the legislation, may ultimately be more problematic than the disease. This paper will discuss some of the details of these provisions along with their explicit and implicit implications not only for financial institutions in the United States, but also for the stakeholders of these institutions both domestic and abroad.

The paper is organized as follows: The next section will discuss the TBTF doctrine which is followed by the provisions in the Dodd-Frank legislation which ostensibly ends the TBTF doctrine. The effects of these provisions on a variety of topics in management theory, specifically the role of a company's board of directors, stakeholder theory and agency theory are then addressed. The implications of this legislation, including the explicit and implicit costs, as well as the potential impact on corporate governance and perhaps the long-term economic growth of the United Sates will follow before concluding with a discussion and recommendations.

2. Too big to fail

Well before the financial crisis of 2008, TBTF was been a topic of discussion with origins traced back to the bank failures and the Great Depression when the United States became acutely aware of the importance of a stable financial system for its economy, and realized the need for some mechanisms to reassure the public to prevent bank runs, and subsequent bank failures, which were common at the time. One mechanism was the creation of the Federal Deposit Insurance Corporation (FDIC) as part of the Glass-Steagall Act of 1933 providing insurance for small, individual deposits up to an initial $2,500 limit and are currently capped at $250,000 (Dodd-Frank Act §335, 2010 p. 165). Also contained within Glass-Steagall was explicit separation of investment banking operations from commercial banking activities.

The Glass-Steagall Act was passed to reassure the American public that monies deposited in a commercial bank would be there even in the event of severe future catastrophic economic downturns. The idea to insure individual deposits and the provisions to separate investment banking and commercial banking was also designed to help mitigate the moral hazard that might otherwise exist by, 1) reassuring and protecting the small savers who offer the foundation for a country's financial system and, 2) by preventing bank holding companies from taking deposits from their commercial banking operations and using them to gain higher, and riskier, returns from investment banking operations.

Following the Great Depression, Glass-Steagall and its tenants seemed to work very well as no significant crises tested the American financial system. This changed, however, in 1984 as a result of the failure of Continental Illinois Bank, then the seventh largest bank in the United States, and the largest bank failure prior to Washington Mutual in the 2008 financial crisis (e.g. Dash and Sorkin, 2008). The decision of the FDIC to recapitalize the bank instead of performing its legal obligation to reimburse small depositors for their loss (at that time $100,000) was seen by some as a critical event (Todd and Thompson, 1990; Stern and Feldman, 2009). Precedent for protecting institutions considered to be vital to the stability of the banking and financial system was set, and reiterated in the epigraph, as some suggest that these institutions are not stand-alone entities, but that a certain level of "interbank" exposure exists between these institutions (see Todd and Thompson, 1990).

Part of the Continental Illinois Bank recapitalization was the saving and loan crisis which, partially as a result of legislation allowing saving and loan institutions to broaden their scope (The S&L Crisis: A Chrono-Bibliography, n.d.) , led to the failure of hundreds of institutions and ultimately led to a taxpayer bailout in 1989 with the Financial Institutions Reform, Recovery and Enforcement Act in an amount just short of $90 billion (General Accounting Office, 1996).

Despite this crisis, a key provision of the Glass-Steagall Act was repealed just 10 years later. Specifically, the Gramm-Leach-Bliley Act of 1999 removed the barrier prohibiting bank holding companies from performing both investment and commercial banking operations in an effort to increase investment yield. At this time, legislation was now in place to both protect the country's financial system and allow more freedom for bank holding companies to conduct their operations which created an environment that both fostered excessively large institutions and contributed to moral hazard. According to Lo (2009), Gramm-Leach Bliley led to an "over-extended financial system" that "could not be sustained indefinitely" (p. 15) and ultimately contributed to the financial crisis of 2008.

This legislation also added to the costs of a bailout when things went awry. Public discontent at the subsequent bailout under the Emergency Economic Stabilization Act of 2008 authorizing the Treasury Department of the United States to spend up to $700 billion, almost seven times larger than the savings and loan bailout, for the purchase of distressed assets under what has become known as the Troubled Asset Relief Program (TARP), prompted Congress explicitly to turn its attention to the TBTF doctrine.

3. Restoring American Financial Stability Act of 2010

After the immediate crisis had passed, experts from varied backgrounds and interests posited that companies may engage in excessive expansion, regardless of the quality of the underlying assets, in an effort to become TBTF (Hughes and Mester, 1992). This may partly explain banks' rapid expansion and consolidation following the 1999 repeal of the Glass-Steagall Act as managers of institutions who believe they have achieved TBTF status may be more likely to engage in excessively-risky behavior knowing that there is some 'backstop' behind their actions should it become necessary. Furthermore, evidence exists to support the notion that creditors to these TBTF institutions believes their "protected" status provided an additional level of confidence to further support, and even encourage, this risky behavior understanding that downside risk is truncated thanks to expected government intervention if necessary (Stern and Feldman, 2009). For example, the insurance provider American International Group (AIG) was effectively nationalized with an $85 billion loan from the federal government (Barr, 2008) and Bank of America received unprecedented pressure from the likes of Treasury Secretary Hank Paulson to merge with Merrill Lynch which was days away from liquidation resulting in substantial losses for Bank of America shareholders (Harris and Kotrozo, 2010).

In response, the Restoring American Financial Stability Act of 2010, also known as the Dodd-Frank Wall Street Reform and Consumer Protection Act, sponsored by Senators Christopher Dodd of Connecticut and Barney Frank of Massachusetts and signed into law on July 21, 2010 by President Obama, is comprehensive reform legislation with numerous purposes. Its more than 800 pages attempt to curtail the agency problems and moral hazard (Jensen and Meckling, 1976; Eisenhardt, 1989) inherent in financial institutions and effectively puts a stop to taxpayer bailouts of these institutions and ends TBTF. (See Kotrozo, 2011, for a more detailed discussion of other provisions found in the legislation.)

Specifically, Title II of the Dodd-Frank legislation outlines the Orderly Liquidation Authority which affects bank holding companies or the majority of companies whose primary Standard Industrial Classification is between 6000 and 6799 (North American Industry Classification, NAICS, codes beginning with 52). This is accomplished, in part, due to the creation of the Financial Stability Oversight Council and the Office of Financial Research (Dodd-Frank Act §112, 2010 pp. 19-20).

The Financial Stability Oversight Council (FSOC), whose duties include the identification of financial stability risks and the promotion of market discipline, consists of 10 voting members and five non-voting members none of which are representatives from the country's financial services institutions themselves (Financial Stability Oversight Council, 2016).

It is supported by the newly-created Office of Financial Research (OFR) which is to, "improve the quality of the financial data and provide analytical support to the FSOC and its members by conducting financial analysis in support of the FSOC and its member agencies and standardizing financial reporting requirements" among other duties (Office of Financial Stability FAQ, n.d. p. 1).

Under section 203(b) of the legislation, a financial company may be put into receivership under the FDIC if the Secretary of the Treasury determines a financial company is, "in default or in danger of default" and, "no viable private sector alternative is available to prevent the default of the financial company" (Dodd-Frank Act §203, 2010 p. 76). This receivership requires approval by both the Board of Governors of the Federal Reserve and a two-thirds vote of the board of directors of the FDIC.

The board of directors of covered financial companies, as those recommended for receivership are referred to in the legislation, may choose to agree with the recommendation of receivership. However, if the financial institution's board of directors does not approve receivership, the Secretary of the Treasury will request an order to do so from the U.S. District Court in Washington, D.C. (Dodd-Frank Act §202, 2010).

Section 204(a) of the bill explicitly addresses moral hazard: PURPOSE OF ORDERLY LIQUIDATION

AUTHORITY.—It is the purpose of this title to provide the necessary authority to liquidate failing financial companies that pose a significant risk to the financial stability of the United States in a manner that mitigates such risk and *minimizes moral hazard* (Dodd-Frank Act §204, 2010 p. 79, emphasis added). The section goes on to say that both creditors and shareholders are to bear the losses of covered financial companies. Further, directors, managers and others who are responsible for the condition of the covered financial company will, "bear losses consistent with their responsibility, including actions for damages, restitution, and recoupment of compensation and other gains not compatible with such responsibility" (Dodd-Frank Act §204, 2010 pp. 79-80). In line with the Sarbanes-Oxley Act of 2002, stewards (directors and officers) of financial institutions can now be held financially liable for mismanagement of financial companies which includes, but is not limited to, incidents of fraud. A specific provision is given that prevents board members from being held liable by shareholders or creditors for a good-faith agreement to receivership, however (Dodd-Frank Act §207, 2010 p. 84).

The FDIC, in liquidating covered financial companies, are mandated that: Such liquidity is necessary to maintain stability of the U.S. financial system, shareholders receive no disbursements until all other claims are paid in full and management and the board of directors of a covered financial company are removed (if such action has not already taken place). The FDIC is also prohibited from taking an equity stake in any covered financial institution (Quintanillo, Albrecht & Brown, 2011), as happened in the AIG bailout, among others.

Furthermore, the bill explicitly states that no taxpayer funds may be used to prevent liquidation of a covered financial company and that taxpayers are not to bear losses from a company placed in receivership and ultimately liquidated (Dodd-Frank §214, 2010).

Finally, the bill also provides the FDIC with new powers. In particular, the bill specifically states that the FDIC may merge the distressed company with another company, or may transfer the covered financial company's assets or liabilities, "*without obtaining any approval, assignment or consent with respect to such transfer*" (Dodd-Frank Act §210, 2010 pp. 86, emphasis added).

4. The implications of Dodd-Frank on corporate governance

Beyond the additional costs and redundancy posed by the new FSOC and OFR, the other provisions of the Dodd-Frank bill intended to put an end to TBTF have potentially-broad implications for corporate governance.

4.1. Weakening the role of a company's board of directors

As a result of the United States' adoption of a system of separation between ownership (shareholders) and control (managers) in publicly-traded companies, shareholders rely on boards of directors who are elected to represent their interest and oversee a firm's management. According to Colley, Doyle, Logan and Stettinius (2005), a company's board of directors are tasked to execute five duties on behalf of shareholders: Fiduciary duty, a duty of loyalty and fair dealing, a duty to care, a duty not to entrench and a duty of supervision, where the board's fiduciary duty is codified in the country's legal system (e.g. Huse and Rindova, 2001).

Rooted in Friedman's (1962) argument that the purpose of a firm is to maximize shareholder wealth within the rule of law, directors' fiduciary duty compels them to consider the monetary and financial interests of shareholders in all of management's dealings, particularly with respect to the board's function to review and, where appropriate, approve the financial objectives, major strategies and plans of the corporation (American Law Institute, 1994; Monks and Minow, 2008).

Unlike most industries, however, an expansion of fiduciary duty extends to the bank's depositors for bank holding companies (Cieri, Sullivan & Lennox, 1994). (Refer to Macey and O'Hara, 2003, for these duties as they relate to banks.) As Adams and Mehran (2003) note, additional expectations are placed on the boards of bank holding companies to reflect a common interest in a country's stable financial institutions. These responsibilities are often found in laws and regulations set forth by a variety of oversight bodies, including Congress. However, corporate governance in financial institutions in general, and bank holding companies in particular, continues to be to maximize shareholder value subject to additional constraints, such as an expanded fiduciary responsibility and greater regulatory oversight.

There is a considerable literature to suggest that banks and a country's financial system contribute to a country's growth (e.g. Levine, 1991; Holmstrom & Tirole, 1993; Levine, 1997; Levine & Zervos, 1998). As such, it is in the best interest of countries to ensure that their financial systems are stable and that additional measures are taken to protect that stability.

In an effort to maintain confidence and stability in the banking sector, the FDIC has always had broad powers with respect to struggling financial institutions. However, the severity of the 2008 crisis, subsequent bailouts and federal support left the American public and its elected officials dissatisfied with having to rescue entities it perceived as having taken unnecessary and unwarranted risks.

It is with this in mind that Treasury Secretary Paulson took an active role in the facilitation of the merger between Bank of America and Merrill Lynch (Harris and Kotrozo, 2010). Subsequently, one of the more significant pieces of the Dodd-Frank legislation is the ability for the Federal Government to put into receivership financial services firms, "in default, or in danger of default" (Dodd-Frank Act §203, 2010 p. 76).

The FDIC's role as the administrator of failing and failed banks has a long history dating back to the Glass-Steagall Act of 1933 and the subsequent Federal Deposit Insurance Act of 1950. Dodd-Frank expands the authority of the FDIC, in conjunction with the Board of Governors of the Federal Reserve, to assume control of financial services firms who are determined to be in danger of default. Boards of directors of such distressed companies, should they disagree with the Federal Reserve Board and FDIC's decision may be the subject of legal action taken by the Secretary of the Treasury in U.S. District Court to compel this action. Furthermore, as previously discussed, Dodd-Frank authorizes the FDIC to merge financial institutions in receivership (also known as "Covered Financial Companies") with another company without that company's approval (Dodd-Frank Act §210, 2010).

The Garn-St. Germain Depository Institutions Act of 1982 granted the FDIC the power to make loans to, deposits in, asset purchases of or contributions to banks covered by FDIC insurance to prevent the closing of the institution or to restore a closed bank. The act also allows the FDIC to facilitate mergers and consolidations via the above instruments to, "guarantee it or the company which controls or will acquire control of it against loss by reason of merging or consolidating" (Garn-St. Germain Depository Institutions Act Bill Summary & Status, n.d.).

In the case of the Bank of America – Merrill Lynch merger, however, the FDIC did not guarantee Bank of America shareholders against the all losses of the merger where these losses rose to $22 billion just days before the consummation of the merger on January 1, 2009 (Paulson, 2010). Instead, the first $10 billion of losses were to be absorbed by Bank of America with additional losses split 90/10 between the government and Bank of America (Paulson, 2010). As such, from the date of the merger announcement between Bank of America and Merrill to January 16th, 2009, Bank of America common shares lost 78% of their value leading the *Wall Street Journal* to characterize it as, "a $50 billion deal from hell" (Moore, 2009). Although directors must also consider its bank's depositors as part of their fiduciary duty, there is no precedent for one bank's board to owe a fiduciary responsibility to the depositors of another financial institution.

As a result, in a departure from the tenants of the Garn-St. Germain Depository Institutions Act of 1982 to use the resources of the FDIC to guarantee Bank of America depositors in the event of a default, Bank of America shareholders were asked to absorb the substantial losses of Merrill Lynch. In effect, the Federal Government split these losses asking shareholders of a sound and healthy financial institution to absorb the "toxic" assets of an institution imminent likely to fail – a practice subsequently codified into law when Dodd-Frank was signed.

Although strong financial stability is crucial to a country's economic growth and development, so too are strong corporate governance mechanisms, such as the roles and functions performed by public companies' boards of directors (Shleifer & Vishney, 1997 and La Porta, Lopez-de-Silanes, Shleifer & Vishney, 2000). Failing to enforce, or weakening, these mechanisms can have an adverse impact on an economy over time, as will be discussed in greater detail shortly.

The perpetual balancing act between protecting the rights of shareholders, which has a long history in the United States, and maintaining stability in the country's financial system, has seen more weight put on the latter with the passage of Dodd-Frank. Proponents of the legislation would suggest that Dodd-Frank levels the scales which have been thrown out of balance, perhaps as a result of the repeal of Glass-Steagall. However, there is research to suggest that corporate governance mechanisms already in place to protect investor interests may be more important. Specifically, Levine (2004) notes, "that it is important to strengthen the ability and incentives of private investors to exert governance over banks rather than relying excessively on government regulators" (p. 1) and, by association, regulations.

In other words, although it is important to protect both the shareholders of banks as well as the other stakeholders who benefit from a country's strong financial system, it may be preferable to focus on what the investors can do to strengthen their investments which will then lead to stronger financial institutions and financial systems.

4.2. Strengthening Stakeholder Theory

By introducing legislation which can supersede the fiduciary duty of a company's board of directors, Dodd-Frank continues a shift away from the value-maximizing tenants of the organization (Friedman, 1962) toward a greater emphasis on Stakeholder Theory (Donaldson and Preston, 1995).

The stakeholders of a country's financial institutions are both numerous and widespread with a strong argument to be made that depositors are the most important among them. Though not necessarily owners of financial institutions, depositors are the foundation of the financial service sector's commercial operations. Simplistically, by helping those with excess capital find those with capital needs, all parties benefit as depositors earn interest on their deposits, those with capital needs pay interest on their loans and the banks collect the difference between these interest rates as profits. The functioning of this relationship between depositors, banks and borrowers has been shown to have a consistent and positive link with a country's economic growth (e.g. Levine, 1997).

Of course, for this relationship to work properly, depositors must have confidence that their monies will be available for withdrawal on demand. To reinforce this confidence, reserve requirements and deposit insurance have been in place and have become widely accepted.

In addition to depositors, numerous other stakeholders, both direct and indirect, exist. Banks' employees and the communities in which the banks are located, competitor and consumer advocacy groups (Adams and Mehran, 2003), the federal government and, indeed, all who rely on the soundness of a country's financial stability. Great pains are undertaken to ensure the confidence and stability in a country's financial institution is maintained. For example, all U.S. taxpayers are stakeholders as tax revenues are used, in part, to guarantee deposits through the FDIC. Although this paper is focused on the U.S. financial services sector, Europe too continues to grapple with its own issues as they relate to Greece, Ireland, Italy, Portugal and Spain, among others having recently created the European Financial Stability Facility (EFSF) whose objective is "to preserve financial stability of Europe's monetary union" (EFSF, 2012 p. 1).

The challenges facing stakeholder theory, however, is a question of both depth and scope. The stakeholder literature is unclear regarding which stakeholder is most important, or put another way, how stakeholders should be subordinated. Determination of the magnitude of importance is also an open question. Consider that the Dodd-Frank legislation may, intentionally or otherwise, put a disproportional burden on the shareholders of these financial institutions. Most stakeholders are somehow compensated for their involvement with an organization: Employees receive a salary, governments receive taxes and depositors receive an interest rate. By comparison, shareholders are entitled to the residual claims on a corporation (Adams, Licht and Sagiv, 2011).

For financial services firm stakeholders who are not somehow directly compensated, though are indirectly compensated through economic growth such as the U.S. or global economy, shareholders of healthy financial service firms may now be asked to bear the risk for financial institutions "that pose a significant risk to the financial stability of the United States" (Dodd-Frank Act §204, 2010 p. 79).

There are many ways these stakeholders could share in the burden of this risk. For example, banks may be asked to pay depositors a higher interest rate for their deposits (which would almost certainly lead to higher borrowing rates), or be required to hold more capital (as is suggested by the Basel Committee as part of Basel III which is also taking place as of 2016), banks could be charged higher taxes, taxpayers could insure a greater level of deposits through the FDIC (which was recently increased from $100,000 to $250,000), as examples. However, with the exception of making the then-temporary increase in FDIC insurance (Gray, 2008) permanent, this legislation ultimately shifts this risk to shareholders.

Events surrounding the merger between Bank of America and Merrill Lynch support this (Harris and Kotrozo, 2010). The exchange between Treasury Secretary Paulson and Bank of America CEO Lewis quoted in the introduction indicates that the Treasury Secretary placed the needs of other stakeholders ahead of those of the shareholders of Bank of America, at the time perhaps in violation of the Garn-St. Germain Depository Institutions Act of 1982. That Dodd-Frank subsequently codifies these actions and allows them to be used again in the future indicates greater risk for investors in U.S. financial institutions.

4.3. Another principle agent

Corporate governance seeks to align the interest of dispersed owners with those of the managers of an organization who may be prone to act in their own best interest as first articulated by Jensen and Meckling's (1976) discussion of agency theory. The two primary mechanisms for aligning these interests are incentives and monitoring. According to Jensen and Meckling (1976) and Tosi, Katz and Gomez-Mejia (1997), monitoring consists of observation of effort through supervision or other devices.

With respect to financial service firms, however, there is the added complexity to the principle-agent relationship which typically involves the management of a firm and that firm's shareholders. Given the importance of other stakeholders, already discussed, Dodd-Frank starts a "tug-of-war" between shareholders, depositors, the board of directors and the financial service firm's management.

In financial service firms, boards of directors (the principles) will monitor management (the agents) to ensure they are acting in the best interest of shareholders, performing its fiduciary duty to maximize shareholders' wealth. However, in doing so, moral hazard is more prevalent, as management may seek to maximize shareholder wealth by taking additional risks with depositors' monies in an effort to gain additional returns.

As such, the federal government steps in and acts as the principle who sees itself as monitoring management (who remain the agent) to ensure it is acting in the best interest of depositors who are the foundation of a country's financial services sector. This leaves a bank's board of directors and its shareholders out of the equation, or at best minimized, offering further support that the federal government perceives stakeholder theory to be more valuable than protecting shareholder rights.

This relationship is not new, of course. Dodd-Frank continues the trend of marginalizing the role of financial service firms' board of directors, however. By infusing the FDIC with powers to place troubled financial firms into receivership, and ultimately merging them with a healthy financial services firm, the Principle-Agent relationship shifts closer to Federal Government acting as the Principle overseeing management with depositors as the primary reason for monitoring activities, with financial stability of paramount importance, instead of the owners (shareholders) of the firm. Boards of directors and shareholders both see their role and influence diminished as a result of the provisions found within this legislation.

5. Beyond and end to TBTF

The effects resulting from the passage of the Dodd-Frank legislation are already being realized, but are likely to extend to the longer-term as well. Both the more immediate and more distant implications are discussed in turn.

5.1. Near-term implications

In an effort to stem moral hazard and prevent a recurrence of the events that lead to the 2008 financial crisis, it is expected that the provisions laid out in the Dodd-Frank legislation, including the creation of the FSOC and OFR, are likely to accomplish the objectives of ending TBTF, mitigating moral hazard and generally, "promot[ing] the financial stability of the United States" (Dodd-Frank Act, 2010 p. 1).

However, these provisions amount to additional monitoring mandates thus leading to higher costs (Jensen & Meckling, 1976). For example, the cost of doing business in the financial services industry will rise as the law allows the OFR to demand "all data necessary" to perform its mandate under the threat of subpoena from the OFR's director for companies which fail to comply. Financial services firms have also expressed concern over the potential dissemination of proprietary information as a result of OFR demands (Schmidt, 2010) as well as the costs and challenges posed by the standardization of financial reporting requirements which allows the OFR to more efficiently conduct its inquiries.

Explicit costs will result from compliance with, defense from and the influence of the specific provisions laid out in the Dodd-Frank legislation, as well as those coming from the newly-created Financial Stability Oversight Council and the Office of Financial Research. While some of the costs have already been realized, others have yet to be quantified or even identified and these costs are not insignificant. Bank of America, for example, posted a $7.3 billion loss due, in part, to Dodd-Frank (Rexrode, 2010). Financial services firms may seek to recoup some of these costs by passing them through to their customers making banking less consumer friendly and more costly (Zywicki, 2011) perhaps leading depositors to seek other alternatives.

It is still too early to tell what the impact of these costs have been. However, Sarbanes-Oxley, the last major piece of financial reform legislation may offer some insights. The Securities and Exchange commission initially estimated the cost of compliance with Section 404 of the Sarbanes-Oxley Act, which deals with a company's internal assessment as well as an outside auditor's assessment of its financial statements, to be $1.24 billion, or roughly $91,000 per company with actual costs estimated as high as $35.0 billion, or $2.87 million per company (Wang, 2008; SEC, 2011; Rosen; 2011).

Implicit costs too will also be incurred as financial services firms move away from their day-to-day operations and strategize about how to deal with the legislation. Additionally, the likelihood of redundancy and inefficiency runs high as the members of the FSOC include the Secretary of the Treasury, Chairperson of the Federal Reserve, the Comptroller of the Currency, Chairperson of the Securities and Exchange Commission and the Director of the Bureau of Consumer and Financial Protection, all Federal entities to which financial service firms were subject to intense scrutiny prior to Dodd-Frank.

One can reasonably argue that these costs are small in comparison to the almost incalculable costs which would be experienced in the event of a collapse of the U.S. financial system, as almost happened in 2008. While these higher costs may be viewed as the "price" of more sound and stable financial system, Dodd-Frank has the potential for additional long-term, implications for financial services firms in general, bank holding companies in particular, and a wide variety of stakeholders in these firms. The next sections will elaborate on the possible, and perhaps unintended consequences of the legislation.

5.2. Long-term implications

The longer-term effects of this legislation will not be felt for some time. However, some lessons from Sarbanes-Oxley, more than 10 years after passage, offer insights here. As recently as 2012, politicians across the political spectrum have expressed an openness to repeal as the burden on business, particularly small businesses, has been substantial. Although the Dodd-Frank legislation itself attempts to mitigate these costs for smaller firms (Dodd-Frank §989, 2010), it is unclear what rules, regulations and requirements are yet forthcoming from the various entities created.

Just as the formation of the Public Company Accounting Oversight Board (PCOAB) under Sarbanes-Oxley created costs and uncertainty for companies, so too does the creation of the Financial Stability Oversight Council and the Office of Financial Research as part of Dodd-Frank. Rules continue to be demanded and crafted as part of the legislation leading to apprehension by financial services firms.

The broad and "unusually strong" powers of the FSOC and OFR, has been referred to as the "CIA of financial regulators" by lobbying groups (Schmidt, 2010), where the law allows the OFR to demand "all data necessary" to perform its mandate under the threat of subpoena from the OFR's director for companies which fail to comply.

In the ever-contentious political environment, there is also the potential for decisions which are politically motivated, though the law's requirement that annual reports by both the FSOC and OFR to Congress may diffuse the potential impact (Dodd-Frank, §112 and §154). However, as Thomas Hoening, President of the Federal Reserve Bank of Kansas City notes, "the final decision of solvency is not market driven but rests with different regulatory agencies and finally with the Secretary of the Treasury, which will bring political considerations into what should be a financial determination" (2011, p. 9).

A significant effect of this legislation, and the one which has received the least attention, is the potential long-term effects this legislation can have on U.S. investment and growth. Specifically, the weakening of investor protection can lead to investment capital to flow out of the United States into countries with stronger shareholder protections as a result of this legislation over the long-term (or the near-term should a crisis again present itself), dampening economic growth which is well established in the literature (Castro, Clementi and MacDonald, 2004; La Porta, Lopez-de-Silanes, Shleifer and Vishney, 2000). Although the United Sates has consistently ranked near the top of countries offering strong corporate governance mechanisms and investor protection (La Porta, Lopez-de-Silanes, Shleifer and Vishney, 1998; Leuz, Nanda and Wysocki, 2003), Dodd-Frank may effectively weaken such protections to the detriment of U.S. competitiveness abroad. Furthermore, shareholders, now realizing they may be required to merge with a distressed company, will begin to require higher rates of return on their financial services investments to compensate for higher levels of risk, which banks will find a hard time producing as a result of the increased monitoring.

Once again consider Sarbanes-Oxley. There is evidence to suggest that the relative competitiveness of the United States in capital market competitiveness has declined considerably since passage in 2002 (Bainbridge, 2011). Somewhat ironically, in what has been known as the Paulson Commission, "one important factor contributing to this trend [the decline in U.S. capital market competitiveness] is the growth of U.S. regulatory compliance costs and liability risks compared to other developed and respected market centers" (Bainbridge, 2010 p. 6).

There is also evidence to suggest that, in light of the costs associated with compliance with Sarbanes-Oxley, the initial public offering (IPO) market has been hurt as private firms look to exit through IPOs less often (Bova, Munitti-Meza, Richardson & Vyas, 2011), and public companies may be going private in an effort to circumvent Sarbanes-Oxley compliance (Engel, Hayes & Wang, 2005; Carney, 2006). These effects and trends will continue as the explicit provisions of Dodd-Frank continue to evolve.

It is also reasonable to believe that this legislation may lead to the exploration and creation of new corporate governance mechanisms to help protect and strengthen managerial control similar to golden parachutes and poison pills found to protect managers from changes in corporate control (Gompers, Ishii & Metric, 2003). Discussion, creation and enforcement of these provisions will take financial service companies away from their primary purpose of facilitating the exchange of capital between those with too much and those with excess need.

A final effect to be felt is the potential for the courts to reexamine a board's fiduciary duty in the case of financial services firms. Will shareholders of firms who are compelled to acquire a financially distressed company have legal regress against the company's board of directors? Although section 207 of Dodd-Frank prevents board members from being held liable by shareholder or creditor for good-faith arrangements in receivership (Dodd-Frank Act §207, 2010 p. 84), this distinction from a board's fiduciary responsibilities will likely be adjudicated. Indeed, the $50 billion class-action lawsuit by shareholders against Bank of America is evidence of this though, admittedly, the lawsuit deals with events prior to the bill's passage (Davidoff, 2011).

6. Discussion

Banking legislation, starting with Glass-Steagall in 1933, including the Savings and Loans crisis in the 1980s, the response to several accounting scandals with Sarbanes-Oxley Act of 2002 and culminating with the Dodd-Frank legislation continues to add layers of complexity to the financial services sector of the United States.

There can be no doubt that oversight is necessary, particularly with such a clear relationship between a country's financial system and its economic growth (Levine, 1997). However, where legislators are elected to craft, debate and enact legislation to benefit both their constituents and the nation as a whole, unintended consequences are often a part of the equation. In addition, there is a tendency to fix errors and omissions of past laws with new legislation adding, instead of subtracting, layers of complexity to the system.

Dodd-Frank seeks to promote the financial stability of the United States, end TBTF and minimize moral hazard, among many other objectives in its 848 pages. To accomplish these objectives, the legislation places additional controls on financial services companies, some of which are still being enacted through the work of the Financial Stability Oversight Council and the Office of Financial Research, two new entities created under the law. The subject of intense scrutiny since the Great Depression early in the 20th Century, banks continue to be a focus of regulators.

Given this attention, one must consider, again, whether all of this regulation and monitoring is having the desired effect. For example, not long after the repeal of Glass-Steagall in 1987, former FDIC Chairman William Isaac stated, "I wonder if we might not be better off today if we had decided to let Continental [Illinois Bank] fail because many of the large banks that I was concerned with might fail have failed anyway. And they probably are costing the FDIC more money by being allowed to continue several more years that they would have had they failed in 1984" (re-quoted in Todd & Thompson, 1990).

Thought of as the foundation on which TBTF was built, the events since the Savings and Loan Crisis of the 1980s leading up to the Financial Crisis of 2008 have some questioning whether the patchwork of laws which both look to protect depositors, the U.S. banking system as a whole, as well as allow banks to increase efficiency, and profitability, by combining both commercial and investment banking activities has become too complex.

In June, 2011, former Federal Reserve Bank of Kansas City President, Thomas Hoenig called for the return of Glass-Steagall regulations to separate investment banking activities from those of commercial banking activities. He refers to Dodd-Frank as an example of the, "good intentions/bad outcomes" syndrome (Hoenig, 2011 p. 3) suggesting the legislation adds complexity to the "tools" already available to address the systemic risk in banking. In speaking about the Federal Deposit Insurance Corporation Improvement Act, another piece of legislation taking aim at TBTF, Stern and Feldman (2009) state, "We do not believe a more formal, more visible role for the Federal Reserve and the Treasury will contribute to a significant change in the incentives that policymakers face when confronted with the bailout decision" (p. 49). One cannot help but wonder whether the same can also be said regarding the aforementioned provisions of Dodd-Frank.

Furthermore as Wayne Abernathy, the Executive Vice-President of the American Bankers' Association notes, "This idea that somehow these guys will notice the economic trends that everyone else is missing, I think that requires a certain suspension of disbelief" (as quoted in Schmidt, 2010).

Was it necessary, in hindsight, for Henry Paulson, then U.S. Secretary of the Treasury, to tell the CEO of Bank of America that the government would not let any systematically important institution fail, and allude to the Treasury's monitoring power to compel Bank of America to consummate their merger with Merrill Lynch? Perhaps, but legislation giving federal regulators the legal authority to put a distressed financial services firm into receivership and then potentially compel the merger between that distressed entity and a healthy entity solves one problem and creates the potential for many more.

Upon announcement of the Bank of America-Merrill Lynch merger, Bank of America shares fell 21% destroying $70 billion in corporate value. Subsequently, Bank of America shares fell almost 90% from mid-2008, before the crisis began, through early March, 2009 devastating shareholders and leading to many changes within the organization, including a shareholder lawsuit which was settled for $2.43 billion in 2012 (Bank of America, 2012).

The Law of Unintended Consequences is ever-present. Legislation spanning 848 pages focused on correcting one problem is more-than-likely going to create many more. It is not yet clear what the overall term effects Dodd-Frank will have on the financial services firms and, perhaps, the U.S. economy as a whole. What is clear, from the bailout of Continental Illinois Bank, to the partial repeal of Glass-Steagall to the passage of Sarbanes-Oxley is that some intended effects were realized along with many other effects which were not. Perhaps this is why former Federal Reserve Bank President Hoening has called for a return to the provisions of Glass-Steagall (Hoenig, 2011), Dodd-Frank itself intends to reduce the costs of Sarbanes-Oxley by exempting some companies as part of Section 989G (Dodd-Frank, 2010 p. 573), and there have been some calls (from conservatives) for the repeal of Dodd-Frank itself (Wyatt, 2011; Johnson, 2010).

Practically speaking, Dodd-Frank is the Law of the Land and is likely to remain so. It is important for lawmakers, from both sides of the political spectrum in a contentious political environment, to take steps to ensure the intended consequences of the legislation are realized and the unintended consequences are avoided, or minimized. For their parts, managers of financial services firms would do well to focus on the spirit of the law in an effort to balance the interests of shareholders, depositors, regulators and the welfare of the economy as a whole. Discussions should take place in Board Rooms and at annual Shareholder Meetings on the best way to minimize the immediate costs of the legislation and strategize about how the legislation will evolve over time, all without significantly detracting from the primary role financial services firms play, simultaneously while understanding the importance their institutions play in the U.S. and global economies.

In conclusion, the most significant piece of financial reform legislation since Sarbanes-Oxley has implications for researchers across a broad range of literatures. The purpose of this paper is not to offer an exhaustive look at the implications for each of these rich research streams such as corporate governance, agency theory and stakeholder theory, but to motivate additional discussion and encourage others to consider how, for example, Dodd-Frank may impact decision-making in Upper Echelon theory (Hambrick and Mason, 1984).

Practitioners too have a vested interest in the evolving nature of this legislation. Ways addressing how boards can continue to emphasize their fiduciary duty to shareholders, or how stakeholders can be accommodated without shifting too much burden onto shareholders are fruitful areas of future research.

Ultimately, this paper poses many more questions than answers. However, clearly articulating some of the questions yet to be answered can help motivate researchers, practitioners and policy makers to ensure that the end of too big to fail is not the start of something worse.

7. References

Adams, R. & Mehran, H. (2003). Is corporate governance different for bank holding companies? *Economic Policy Review*, Federal Reserve Bank of New York, April, 123-142.

Adams, R. B., Licht, A. N. & Sagiv, L. (2011). Shareholders and stakeholders: How do directors decide? *Strategic Management Journal, 32, 1331-1355.*

American Law Institute. (1994). *Principles of corporate governance: Analysis and recommendations.* St. Paul, MN: American Law Institute Publishers.

Bachus, S. & Hensarling, J. (2011). One year later: The consequences of the Dodd-Frank Act. Financial Services Committee, 1-30.

Bainbridge, S. M. (2010). Corporate governance and U.S. capital market competitiveness. UCLA Working Paper, 1-36.

Bank of America. (2012). http://newsroom.bankofamerica.com/press-release/corporate-and-financial-news/bank-america-reaches-settlement-merrill-lynch-acquisition accessed March 1, 2016.

Barr, A. (2008). AIG slumps after signing up for $85 billion Fed loan. *Wall Street Journal* Digital Network September 24, http://www.marketwatch.com/story/aig-nationalized-as-insurer-signs-up-for-85-billion-fed-loan, accessed October 15, 2010.

Bova, F, Minutti-Meza, M., Richardson, G. & Vyas, D. (2011). The Sarbanes-Oxley Act and exit strategies of private firms. University of Toronto Working Paper, 1-43.

Carney, W. J. (2006). The cost of being public after Sarbanes-Oxley: The irony of "going private." *Emory Law Journal, 55, 141-160.*

Castro, R., Clementi, G. L. & MacDonald, G. (2004). Investor protection, optimal incentives and economic growth. *Quarterly Journal of Economics, 119, 1131-1175.*

Cieri, R. M., Sullivan, P. F. & Lennox, H. (1994). The fiduciary duties of directors of financially troubled companies. *Journal of Bankruptcy Law and Practice, 3, 405-422.*

Colley, J., Doyle, J. L., Stettinius, W. & Logan, G. (2005). *What is corporate governance?* New York: McGraw Hill/Irwin.

Dash, E. & Sorkin, A. R. (2008). Government seizes WaMu and sells some assets. *New York Times* Online Edition September 25, http://www.nytimes.com/2008/09/26/Business/26wamu.html, accessed August 4, 2011.

Davidoff, S. M. (2011). A $50 billion claim of havoc looms for Bank of America. *Dealbook New York Times*, September 27, http://dealbook.nytimes.com/2011/09/27/for-bank-of-america-a-looming-50-billion-claim-of-havoc/, accessed November 12, 2011.

Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010, Pub. L 111-203 H.R. 4173 2010.

Donaldson, T. & Preston, L. E. (1995). The stakeholder theory of the corporation: Concepts, evidence and implications. *Academy of Management Review, 20, 65-91.*

Eisenhardt, K. M. (1989). Agency Theory: An assessment and review. *Academy of Management Review, 14, 57-74.*

Emergency Economic Stabilization Act of (2008). H.R. 1424 110th Congress: In *GovTrack.us (database of federal legislation).* Retrieved February 4, 2012, from http://www.govtrack.us/congress/bill.xpd?bill=h110-1424.

Engel, E., Hayes, R. M. & Wang, X. (2005). The Sarbanes-Oxley Act and firms' going private decisions. *Journal of Accounting and Economics, 44, 116-145.*

European Financial Stability Facility. (2012). http://www.efsf.europa.eu/attachments/faq_en.pdf, accessed February 4, 2012.

Federal Deposit Insurance Act of 1950, Pub. L 81-797, 64 Stat. 873.

Financial Institution Reform, Recovery and Enforcement Act of 1989, Pub. L 101-73, 103 Stat. 183.

Financial Stability Oversight Council. (2016). https://www.treasury.gov/initiatives/fsoc/about/council/Pages/default.aspx accessed March 10, 2016.

Friedman, M. (1962). *Capitalism and freedom*. Chicago: The University of Chicago Press.

Garn-St. Germain Depository Institutions Act of 1982, Pub. L 97-320, 96 Stat. 1469.

Garn-St. Germain Depository Institutions Act Bill Summary & Status 97th Congress (1981-1982) H.R. 6267 All Information. n.d. *The Library of Congress*, http://thomas.loc.gov/cgi-bin/bdquery/z?d097:HR06267:@@@L&summ2=m&|TOM:/bss/d097query.html, accessed February 4, 2012.

General Accounting Office. (1996). Financial audit: Resolution trust corporation's 1995 and 1994 financial statements. B-262036, 1-52.

Glass-Steagall Act of 1933, Pub. L 73-66, Stat. 162.

Gompers, P., Ishii, J. & Metrick, A. (2003). Corporate governance and equity prices. *Quarterly Journal of Economics, February, 107-155.*

Gramm-Leach-Bliley Act of 1999, Pub. L 106-102, 113 Stat. 1338.

Gray, A. (2008). Emergency Economic Stabilization Act of 2008 temporarily increased basic FDIC insurance coverage from $100,000 to $250,000 per depositor. Federal Deposit Insurance Corporation, October 7, 2008, http://www.fdic.gov/news/news/press/2008/pr08093.html, accessed February 4, 2012.

Hambrick, D. C. & Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review, 9, 193-206.*

Harris, R. & Kotrozo, J. (2010). Bank of America's acquisition of Merrill Lynch: A multi-theoretical analysis. *The Journal of Management and Engineering Integration*, *3, 21-32*.

Hoenig, T. M. (2011). Financial reform: Post crisis? Women in Housing and Finance, Washington, D. C., February 23, 1-13.

Holmstrom, B. & Tirole, J. (1993). Market liquidity and performance monitoring. *Journal of Political Economy, 101, 678-709.*

Hughes, J. P. & Mester, J. (1992). A quality and risk-adjusted cost function for banks: Evidence on the too big to fail doctrine. *Journal of Productivity Analysis, 4, 293-315.*

Huse, M. & Rindova, V. P. (2001). Stakeholders' expectations of board roles: The case of subsidiary boards. *Journal of Management and Governance, 5, 153-178.*

Jensen, M. C. & Meckling, W. H. (1976). Theory of the firm. Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3, 305-360.*

Johnson, S. (2010). Ending 'Too big to fail.' *New York Times* Online Edition April 8, http://economix.blogs.nytimes.com/2010/04/08/ending-too-big-to-fail/ accessed October 8, 2010.

Kotrozo, J. (2011). Making sense of Dodd-Frank from a corporate governance perspective. *Journal of Business Management and Change, 6, 7-28.*

La Porta, R., Lopez-de-Silanes, F., Shleifer, & A., Vishny, R. (1998). Law and finance. *Journal of Political Economy, 106, 1115-1155.*

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. (2000). Investor protection and corporate governance. *Journal of Financial Economics, 58, 3-27*.

Leuz, C., Nanda, D., & Wysocki, P. (2003). Earnings management and investor protection: An international comparison. *Journal of Financial Economics, 69, 505-527.*

Levine, R. (1991). Stock markets, growth and tax policy. *Journal of Finance, 46, 1445-1465.*

Levine, R. (1997). Financial development and economic growth: Views and agenda. *Journal of Economic Literature, 35, 688-726.*

Levine, R. (2004). The corporate governance of banks: A concise discussion of concepts and evidence. *World Bank Policy Research Working Paper, 3104, September, 1-20*.

Levine, R. & Zervos, S. (1998). Stock markets, banks and economic growth. *The American Economic Review, 88, 537-558.*

Lo, A. W. 2009. Regulatory reform in the wake of the financial crisis of 2007-2008. *Journal of Financial Economic Policy, 1, 4-43.*

Macey, J. R. & O'Hara, M. (2003). Corporate governance problems in banks. *Economic Policy Review*, Federal Reserve Bank of New York, 9, April, 91-107.

Mayopolous, T. J. (2009). Statement of Timothy J. Mayopolous Before the United States House of Representatives Committee on Oversight and Government Reform and Subcommittee on Domestic Policy. November, 17, http://democrats.oversight.house.gov/images/stories/Hearings/Committee_on_Oversight/TESTIMONY-Mayopoulos.pdf, accessed February 4, 2012.

Monks, R. & N. Minow (2004). *Corporate governance*. Cambridge, MA: Blackwell Business.

Moore, H. N. (2009). Bank of America-Merrill Lynch: A $50 billion deal from hell. *Wall Street Journal Blogs*, January 22, 2009, http://blogs.wsj.com/deals/2009/01/22/bank-of-america-merrill-lynch-a-50-billion-deal-from-hell, accessed November 2, 2010.

Office of Financial Research Created Under the Dodd-Frank Wall Street Reform and Consumer Protection Act: Frequently Asked Questions. n.d.. ttp://www.treasury.gov/initiatives/Documents/OFR_FAQ-11242010-FINAL.PDF, accessed February 4, 2012.

Paulson, H. (2010). *On the brink: Inside the race to stop the collapse of the global financial system*. New York: Business Plus.

Quintanilla, R., Albrecht, M & Browne, B. (2011). The U.S. Government says support for banks will be different "next time" – but will it? *Standard and Poor's RatingsDirect on the Global Credit Portfolio*, July 12.

Rexrode, C. (2010). $7.6 Billion loss as Bank of America braces for new rules. *Lexington Herald Leader* Online October, 19, http://www.kentucky.com/2010/10/19/1486472/76-billion-loss-as-bank-of-america.html, accessed October 24, 2010.

Rosen, J. 2011. (2012). Candidates target costly Enron-Era law aimed at thwarting accounting fraud. Foxnews.com, http://www.foxnews.com/politics/2011/10/27/decade-old-accounting-regulation-becomes-popular-target-on-2012-campaign-trail/, October 27, accessed November 12, 2011.

Sarbanes-Oxley Act of 2002, Pub. L 107-204, 116 Stat 745 H.R. 3763 2002.

Schmidt, R. (2010). The Treasury's New Research Office. *Bloomberg Businessweek*, September 2, http://www.businessweek.com/magazine/content/10_37/b4194026970590.htm, accessed September 6, 2010.

Securities and Exchange Commission. (2011). Study and Recommendation of the Sarbanes-Oxley Act of 2002 for issuers with public float between $75 and $250 million. April, 1-118.

Shleifer, A. & Vishney, R. W. (1997). A survey of corporate governance. *Journal of Finance, 52, 737-783.*

Stern, G.H. & Feldman, R.J. (2009). *Too big to fail: The hazards of bank bailouts*. Harrisonburg, VA: R.R. Donnelley.

The S&L Crisis: A Chrono-Bibliography. n.d. http://www.fdic.gov/bank/Historical/s&l/, accessed August 4, 2011.

Todd, W. F. & Thompson, J. B. (1990). An insider's view of the political economy of the too big to fail doctrine. *Federal Reserve Bank of Cleveland Working Paper 9017, 1-86.*

Tosi, H. L., Katz, J. P. & Gomez-Mejia, L. R. (1997). Disaggregating the agency contract: The effects of monitoring, incentive alignment and term in office on agent decision making. *Academy of Management Journal, 40, 584-602.*

Wang, J. (2008). Sarbanes-Oxley Section 404 places disproportionate burden on smaller public companies. *The Heritage Foundation Center for Data Analysis, August, 1-40.*

Wyatt, E. (2011). Dodd-Frank Act a favorite targets for Republicans laying blame. *New York Times*, September 20, http://www.nytimes.com/2011/09/21/business/dodd-frank-act-is-a-target-on-gop-campaign-trail.html?pagewanted=all, accessed November 12, 2011.

Zywicki, T. (2011). The Dick Durbin bank fees: Bank branches close while banking becomes more expensive and less consumer-friendly. *Wall Street Journal*, September 29, http://online.wsj.com/article/SB10001424052970204831304576597173130633798.html, accessed December 21, 2010.

Results and Implications of an Engineering Capstone Course Faculty Survey

Harvey F. Hoffman
Fairfield University
hhoffman@fairfield.edu

Abstract

The findings from a survey of engineering and technology faculty show that many faculty members do not have business and industry engineering management experience. Engineering capstone courses frequently focus solely on a project's technical aspects. Faculty may benefit from guidance on the business practices that should be included in an engineering capstone course.

Above and beyond the ability to design and analyze products, services, processes and systems, entry-level engineers must be able to communicate with the appropriate management level in their organization. A failure to satisfactorily address engineering or technology project topics such as need, schedule, budget, cost, risk, return-on-investment, and resources required when presenting a proposal for a new venture may likely doom the effort. A young engineer's project management skills may be the key to early career advancement.

A pedagogical approach to teaching a capstone course in the form of a project template is discussed that simulates an engineering project as conducted in industry.

1. Introduction

Engineering is about designing, manufacturing, testing and selling products and services. Engineers make things better, safer, more reliable, with improved quality and performance while working within the constraints of a specification, budget and schedule, all the while keeping a customer happy. Engineering is a pragmatic profession. Engineers are builders, entrepreneurs, problem solvers and, "in general, make technology work [1]." Engineers in the commercial arena participate in increasing the organization's revenue, lowering operating expenses, and improving profitability. Engineers must "cope with the commercial realities of industrial practice in the modern world, as well as the legal consequences of every professional decision they make [2]."

There exists a tension in many engineering curricula between providing 1) a broad liberal arts education, 2) a technical course of studies that meets the breadth and depth preferred by the engineering faculty, and 3) preparing students to ethically execute in an industry that is profit motivated. A typical baccalaureate-engineering curriculum leaves little time for the nascent engineer to learn about the business aspects of organizations. This may result in a lack of understanding of the importance of their contribution to the company's bottom line, which may hinder their career progress. To be sure, the engineer's work is critical to the success of a technology department or company, but young engineers should not delude themselves - whoever controls the finances controls the decisions made in the organization. Engineers must understand and align their work with the organization's financial goals.

The purpose of this paper is to examine and highlight the idea that in our competitive and entrepreneurial global economy successful engineers need technical competency and business skills. The paper examines the practical business and management background of a group of engineering faculty. A perspective is offered in the form of a project template to guide faculty with minimal business experience to help their students "hit the ground running" when they enter industry.

The results of a survey prepared for this research examines the topics covered by professors teaching engineering capstone or senior project courses, and their business and management background. Suggestions are made to structure an engineering capstone or senior project course report template to assist a faculty member in preparing the budding engineer for the realities of management practices and expectations in industry.

2. Engineers and business practices

National Science Foundation data [3] show that 79% of engineering graduates enter business or industry (Table 1). Another 15% enter the nonprofit or government areas and 6% enter the education field. Whether working for a nonprofit, for-profit or governmental organization, the work recent engineering graduates will engage in will likely be part of a broad strategy that involves both technical and financial objectives.

Table 1. Distribution of Engineering Employment in USA

Engineering Employment Distribution in USA	Number Employed	Percentage
Business or industry	1,173,000	76.3%
Federal Government	105,000	6.8%
Nonprofit	31,000	2.0%
Self-employed	40,000	2.6%
State or Local Government	94,000	6.1%
Universities and 4-year Colleges	88,000	5.7%
Other Educational Institutions	6,000	0.4%
Total	1,537,000	

Source: Data extracted from Women, Minorities, and Persons with Disabilities in Science and Engineering, Arlington, VA. NSF 15-311 January 2015. Table 9-19. Employed scientists and engineers, by sector of employment, broad occupation, sex, ethnicity, race, and disability status: 2013. [3]

Management in industry makes project and business decisions based on product or service need, novelty, cost reduction, safety, return-on-investment, resource availability, societal impact, and intellectual property considerations as well as technical feasibility. It becomes imperative that engineering faculty convey basic business knowledge to students to prepare them to provide answers to the questions that managers ask.

In a limited way, the Fundamentals of Engineering (FE) exam [4] acknowledges the importance of business practices to the new engineer by posing exam questions related to engineering economics, time value of money (e.g., present value, future value, annuities), intellectual property, cost estimation, risk identification, and cost analysis (e.g., cost-benefit, trade-off, breakeven). Depending on the discipline, the January 2014 Fundamentals of Engineering (FE) examination assigned from 3% to 21% of the 110 questions to business practices (Table 2).

The realities of the engineering world dictate that technological skill is only a part of what an engineer needs to succeed. Engineering practice consists of more than the application of science to solve problems. Knowledge of mathematics and engineering theory is necessary but not sufficient to be a successful engineer. The seminal engineering education 1994 Green Report [1] was among the first to recognize that engineers needed more than technical skills to be successful. The report encouraged engineering educational institutions to incorporate a broad framework for engineering curricula that included (among other important criteria)

> An appreciation of different cultures and business practices, and the understanding that the practice of engineering is now global;
> A multi-disciplinary perspective;
> A commitment to quality, timeliness and continuous improvement;
> Undergraduate research and engineering work experience;
> Understanding of the societal, economic and environmental impacts of engineering decisions; and
> Ethics.

Table 2. Fundamentals of Engineering Business Practices

Fundamentals of Engineering Exam	Knowledge Area	No. of Questions in Exam	Maximum Percentage of Business Practice Questions [2]
Chemical	Ethics and Professional Practice	2-3	3%
Civil	Engineering Economics Ethics and Professional Practice	4-6 4-6	11%
Electrical and Computer	Engineering Economics Ethics and Professional Practice	3-5 3-5	9%
Environmental	Engineering Economics Ethics and Professional Practice	4-6 5-8	13%
Industrial	Engineering Economics Ethics and Professional Practice	10–15 5-8	21%
Mechanical	Engineering Economics Ethics and Professional Practice	3-5 3-5	9%
Other Disciplines	Engineering Economics Ethics and Professional Practice	7-11 3-5	15%

The Green report led to the development of the Accreditation Board for Engineering and Technology (ABET) [5] outcomes (a) – (k) criteria in 2000. The ABET revision to criteria 3 requires engineers to make informed judgments, which must consider the impact of engineering solutions in global, economic, environmental, and societal contexts.

It is no longer enough to come out of school with a purely technical education; engineers need to be entrepreneurial in order to understand and contribute in the context of market and business pressures [6]. Attention to intellectual property, project management, multilingual influences and cultural diversity, moral and religious repercussions, global and international impacts, national security, and cost-benefit constraints drive engineering practice [7]. Those engineers who master the principles of business and management will likely be rewarded with leadership roles. With the growing interdependence between technology and the economic and social foundations of modern society, there will be an increasing number of opportunities for engineers to exercise their potential as leaders, not only in business but also in the nonprofit and government sectors [7]. The more an engineer can expand his or her expertise to include knowledge of business operations and practices, the more valuable that engineer will be to the organization.

3. The Capstone Course

The capstone course is intended to be the final course in a major. It makes use of prior coursework and life experience in a cumulative and integrative fashion. An engineering capstone course incorporates all skills and competencies that students have learned in their engineering program. It should balance technical, business, and interpersonal skills that will help students to immediately contribute to a team's efforts. The engineering capstone course simulates as close as academically possible the activities in which an engineer is involved. The course challenges the student's personal and professional skills. The nature of the course forces students to accept dimensions of professional practice that go beyond technology by also including societal considerations. Most engineering capstone courses expect participants to apply the information gained from engineering, English writing, economics, statistics, accounting, management, marketing, law, ethics and computer engineering to the task-at-hand.

Many authors ([8], [9], [10], [11], [12], [13]) have discussed the purpose of the capstone course at the undergraduate and graduate levels. These authors agree that a capstone course should include experiences that benefit the graduate or senior undergraduate student learner by:

- Using technical knowledge towards solving an engineering problem that incorporates "real world" practical issues and challenges.
- Integrating technical and non-technical competencies.
- Using management techniques to mitigate and overcome problems and issues that invariably arise during a project.
- Engaging in a team activity that calls upon the student's to use collaboration, conflict resolution and brainstorming skills.
- Practicing oral and written communication skills.

4. Faculty Survey

A survey conducted in support of this research primarily used the e-mailing listservs for the Engineering Technology Division of the American Society for Engineering Education (ASEE) and the American Society of Mechanical Engineers (ASME). The online survey received 95 responses, 93% of whom were full-time faculty, 59% taught undergraduate capstone courses only and 40% taught both undergraduate and graduate capstone courses. The majority of the respondents teaching capstone courses worked in industry (Table 3). Manufacturing and aerospace were the primary areas in which the respondents worked with mechanical and electrical engineering or engineering technology as the major disciplines (Table 4). The percentages in the table total more than 100% because respondents selected multiple areas. Faculty are consistent in the expectation that capstone students engage in design and analysis, prepare written and oral reports, work in teams, engage in design reviews and test a prototype or proof-of-concept unit (Table 5).

Table 3. Survey Respondents Working in Industry

Years Worked in Industry	Number of Respondents - %
Less than 1 year	12%
From 1 to 3 years	15%
From 4 to 6 years	14%
From 7 to 10 years	17%
More than 10 years	42%

Table 4. Technical Areas in Which Survey Respondents Worked

Technical Area	Area Selected by Respondents - %
Mechanical	33.3%
Manufacturing	29.8%
Electrical	22.6%
Aerospace	17.9%
Electronic	17.9%
Research and development	17.9%
Project management	15.5%
Software	11.9%
Test	11.9%
Computer	10.7%
Systems	10.7%
Structural	8.3%
Civil	7.1%
Process	7.1%
Biomedical	4.8%
Energy	4.8%
Transportation	4.8%
Nuclear	3.6%
Business development	2.4%
Agricultural	1.2%

Table 5. Topics Included in Survey Respondent's Capstone Course

Engineering Capstone Topics	Faculty Including Topics in Capstone Course -%
Design	93.50%
Written report	88.20%
Oral presentation	87.10%
Teamwork	83.90%
Analysis	81.70%
Prototype	71.00%
Testing	62.40%
Budget preparation and monitoring	61.30%
Design review	59.10%
Schedule preparation and monitoring	57.00%
Constructing	54.80%
Drawing preparation	52.70%
Economic factors	51.60%
Functional requirements definition	51.60%
Ethics	50.50%
Safety	50.50%
Specification writing	44.10%
Manufacturing feasibility	43.00%
Simulation	41.90%
Coding	37.60%
Intellectual property	32.30%
Proof of principle	32.30%

Engineering Capstone Topics	Faculty Including Topics in Capstone Course -%
Social impact	28.00%
Environmental impact review	26.90%
Entrepreneurship	25.80%
Human factors	25.80%
Patent review	25.80%
Risk identification and mitigation planning	24.70%
Aesthetics	23.70%
Competitive analysis	21.50%
Return on investment	21.50%
Maintainability	19.40%
Life-cycle cost	17.20%
Reliability	17.20%
Product or service differentiators	10.80%
Availability	7.50%
Sole source issues	5.40%

A manager in business and industry is frequently involved in planning, organizing, delegating, communicating, recruiting personnel, setting and monitoring internal and external standards, setting priorities, allocating resources, examining competitive analyses, and reviewing project status and work accomplished. The number of direct reports a manager has is sometimes referred to as his or her "span of control." Although no perfect ratio likely exists, many authors suggest that the number of a manager's direct reports is between 5 and 10 people ([14] and [15]). The survey results indicated that of the 59% of the respondents who had spent seven or more years in industry only 41% of the respondents managed four or more people. Approximately 80% of the respondents did not have profit/loss responsibility. The capstone course does engineering students a disservice if it does not prepare engineering students to understand the mindset of the engineering manager who has finances at the top of the list of concerns.

5. Project report template

The survey showed that some faculty members do not have experience in business, engineering management or entrepreneurship and do not have a good understanding of these areas [16]. The following project report template evolved from discussions with engineering managers representing several industrial firms. They were members of the department's Industry Advisory Board. The template is offered as a guide for faculty in preparing engineering and technology capstone students to anticipate the questions that a manager may ask about a new project or proposal opportunity.

1. Cover Sheet

Title of Project, Team Members, Course Number, Course Title, Date Submitted

2. Description of the Project Problem

Describe the problem or need that the team is addressing. Identify the purpose or objective of the project, the context of the project and the general technical problem. Describe the tasks to be performed. Include information about the research, analysis, hardware design, software design and cost effectiveness (e.g., manufacturability). Identify required resources and provide a task schedule.

3. Background

Discuss the context and history of the selected topic and describe what has been done in the past. Include the results of a literature and intellectual property search. Other issues that may be addressed include

Economic or financial - effect of this topic on the local economy, savings, material cost, labor issues, outsourcing needs, etc.

Environmental – product influence on the environment over product's life cycle.

Sustainability

Manufacturability - material availability, use of commercial-off-the-shelf (COTS) versus custom components, special needs

Ethicality - potential harm/benefit to society; ethical issues

Health and safety - positive or negative impacts on the health and safety of individuals or society for past or future applications

Social - relationship of this topic to education, culture, communication, entertainment

Political - relationship of this topic to political issues

4. Design Requirements or Project Specification

The detailed design criterion should address goals for performance, reliability, cost, technical risks, manufacturability, safety, and societal factors (human interface, environmental factors, product disposal, etc.). Identify tasks specifically excluded from the scope of this project. The main purpose of the specification is to communicate to all stakeholders the work to be done. A disagreement or misunderstanding is easy to correct early in the process. Discovered later on, work must be done to correct the problem - costing time and money. The specification is the foundation of the project.

5. Test

Describe what constitutes project success and why? Discuss the product or service tests that will confirm that the project succeeds in doing what it was intended to do.

6. Organization or Business Aspects

Discuss why an organization should invest money in this product or service development. Consider addressing the following:

a) Describe the market and/or industry
b) What is the economic outlook for the industry?
c) Describe the novel features of this product or service. Why is this product or service different from or better than the competition's?
d) Competitive landscape – what is out there that is similar to the proposed product or service?
e) Describe the intellectual property or patent issues – if any?
f) Who are the projected customers or clients?
g) Describe the resources needed to develop, test, evaluate, and market the product or service.
h) What are the key competitive factors in this industry?
i) What are the potential risk areas and how would the team mitigate them?
j) What is required to bring the product or service to market?
k) How long will it take to recover the investment made to develop the product or service?
l) What would the financial return be over the first three year time frame after product or service deployment? What assumptions are made in this calculation?

7. Final Implementation and Deliverables - Discuss the prototype realization (including diagrams, photos, videos, code examples, and other technical drawings). Discuss and present the calculations that support critical decisions.
8. Possible Future Work

6. Conclusion

Engineering is not only about applying scientific laws and principles to technical problems. It is part of the fabric of society and it necessarily brings engineers and technologists into the mainstream of organizational decision making. Entry-level technical graduates should be involved in situations that call for an understanding of basic business decision making criteria. The survey conducted for this research indicates that many faculty teaching capstone courses may not have sufficient managerial experience to respond to basic business questions. Indeed, since 1995 Boeing has been working with faculty to improve their understanding of key engineering business realities with the expectation that they will return to the classroom and emphasize practical project based learning [17].

Decision making is as important as technical skill in technical design work. Managers must analyze data, assign priorities, and make project go-no go decisions based on practical matters. The proposed project template is a guide for both faculty and students. The template will not make managers out of new graduates. The hope is that it will assist a new graduate to anticipate many of the questions that a manager may pose about a project so that they will have a satisfactory response prepared in advance. Even in an engineering-dominated organization, gaining a better understanding of the business reasons for a decision will make a recent graduate a better technical person.

The results of the survey and the suggestions for the project report represent a step in understanding, assessing, and improving the engineering capstone course.

7. References

[1] American Society for Engineering Education, "The Green Report - Engineering Education for a Changing World," American Society for Engineering Education, October 1994. [Online]. Available: http://www.asee.org/papers-and-publications/publications/The-Green-Report.pdf. [Accessed 15 May 2015].

[2] J. E. Mills and D. F. Treagust, "Engineering education – is problem based or project-based learning the answer?," 2003. [Online]. Available: http://www.aaee.com.au/journal/2003/mills_treagust03.pdf. [Accessed 20 November 2016].

[3] National Science Foundation: Women, Minorities, and Persons with Disabilities in Science and Engineering, "National Center for Science and Engineering Statistics, Scientists and Engineers Statistical Data System (SESTAT), 2013 (preliminary).," January 2015. [Online]. Available: http://www.nsf.gov/statistics/2015/nsf15311/tables.cfm. [Accessed 15 May 2015].

[4] National Council of Examiners for Engineering and Surveying, "FE Exam," 2015. [Online]. Available: http://ncees.org/exams/fe-exam/. [Accessed 26 May 2015].

[5] Accreditation Board for Engineering and Technology (ABET), 2014 - 2015 Criteria for Accrediting Engineering Programs, Baltimore: ABET, 2013.

[6] T. Byers, T. Seelig, S. Sheppard and P. Wellerstein, "Entrepreneurship: Its Role in Engineering Education," *Summer issue of The Bridge on Undergraduate Engineering Education,* vol. 43, no. 2, pp. 35-40, Summer 2013.

[7] National Academy of Engineering, "The Engineer of 2020: Visions of Engineering in the New Century," National Academies Press, Washington, 2004.

[8] A. S. Thomas, "The business policy course: Multiple methods for multiple goals," *Journal of Management Education,* vol. 22, pp. 484-497, 1988.

[9] D. Moore and F. Berry, "Industrial Sponsored Design Projects Addressed by Student Design Teams," *Journal of Engineering Education,* vol. 90, no. 3, pp. 69-73, July 2001.

[10] R. Pimmel, "Cooperative Learning Instructional Activities in a Capstone Design Course," *Journal of Engineering Education,* vol. 90, no. 3, pp. 413-421, July 2001.

[11] E. Grubbs and M. W. Ostheimer, "Real World Capstone Design Course," in *Proceedings of the 2001 American Society for Engineering Education Annual Conference & Exposition - Session 3232*, 2001.

[12] D. W. Dinehart and S. P. Gross, "A service learning structural engineering capstone course and the assessment of technical and non-technical objectives," *Advances in Engineering Education,* vol. 2, no. 1, Spring 2010.

[13] H. F. Hoffman, The Engineering Capstone Course: Fundamentals for Students and Instructors, New York: Springer, 2014.

[14] R. Ashkenas, "More Direct Reports Make Life Easier," 25 September 2012. [Online]. Available: https://hbr.org/2012/09/more-direct-reports-make-life.html .

[15] G. L. Neilson and J. Wulf, "How Many Direct Reports?," *Harvard Business Review,* pp. Vol 90, no. 4, April 2012.

[16] S. Zappe, K. Hochstedt, E. Kisenwether and A. Shartrand, "Teaching to innovate: Beliefs and perceptions of instructors who teach entrepreneurship to engineering students," *The International journal of engineering education,* pp. 45-62, 2013.

[17] K. V. Treuren, D. Kirk, T-min. Tan and S. Santhanam, "Developing The Aerospace Workforce: A Boeing Experience," in *American Society for Engineering Education Annual Conference & Exposition*, 2010.

Economic Approach for Determining the Time Buffer in Synchronous Manufacturing

Faisal S. AlGhamdi and Gamal S. Weheba
Department of Industrial and Manufacturing Engineering
Wichita State University
Wichita, KS 67260 USA
gamal.weheba@wichita.edu

Abstract

There is a growing need for models that can determine the time buffer in applications of Theory of Constraints (TOC) for production control. In particular, applications of the Drum Buffer Rope (DBR) and the Simplified Drum Buffer Rope (SDBR). In this research, a new model was developed for determining the size of the time buffer based on six process and cost parameters. The model was developed based on the Taguchi's Loss Functions and allows a tradeoff between capacity and delay losses in determining the buffer size. This provided means for overcoming limitations of existing models while maintaining simplicity as an appealing feature for practitioners. A numerical example is included to illustrate the model applicability and highlight its advantages. In addition, statistical analysis was conducted to investigate model performance over a wide range of its parameters. The results indicated that the model is most sensitive to changes in the processing time at the CCR and the maximum delay that can be tolerated.

1. Introduction

Synchronous manufacturing is a production technique that synchronizes the entire process based on the system's bottleneck or constraint. It coordinates all resources together in the plant. This harmonizing logic (synchronized materials) was early introduced by Goldratt and Cox (1984). The notion behind this concept is to efficiently utilize all resources together and to improve the system's performance. This logic is also known as theory of constraints, or simply TOC (Frazier & Reyes, 2000). Under the Synchronous manufacturing philosophy, the physical inventory or work in process (WIP) is eliminated by establishing a time buffer. The time buffer prevents the accumulation of WIP in front of the bottleneck (CCR), or reduces it to a minimal level. The focus is on the system's flow rather than capacity. It is crucial to manage the system's constraint wisely and effectively to improve the system's throughput and to make money (Sivasubramanian et al., 2003).

Theory of Constraints (TOC) is a management philosophy developed by Goldratt in the late 1970s which assumes that there is always at least one constraint inherent in any system. However, this constraint is considered the weakest link in the process. The concepts have evolved steadily since their introduction. For example, Watson et al. (2007); Tulasi and Rao (2012); Rahman (1998), Cox and Schleier (2010) and many others provided reviews and applications of TOC in production control. These concepts were also applied in the service sector including healthcare and banking. Examples include Siha (1999) Robbins (2011), Groop (2012), Sadat et al. (2013), and Peltokorpia et al. (2016).

The essence of the TOC lies in identifying the constraint which is defined as "Anything that limits a system from achieving higher performance versus its goal" Goldratt (1988). According to Dettmer (2000), the type of constraint is either physical or policy (non-physical). All labor, machines, tools and any tangible resources that can limit the system's performance are physical constraints. However, the policy constraints are about managerial practices that can slow the performance of the system and affect its productivity (i.e. market segments).

As indicated by Spencer and Cox (1995), TOC can be divided into three different categories: (1) Logistics, (2) Performance System (Performance Measurement), and (3) Problem Solving/Thinking Process. The preliminary part of the logistics category is the the Drum Buffer Rope (DBR) for application in production control. It is derived from the five focusing steps (5FS) and evolved to a new generation known as simplified-drum-buffer-rope (SDBR). A production control technique that treats the market as the true constraint for the entire system. All operations must be subordinated to the market even if the internal capacity is constrained. This concept elevates the system's capacity to the level of market demands despite internal restrictions. The preliminary aspect of SDBR is guaranteeing safe product delivery by ensuring the due date is less than the quoted lead time (QLT), which is assumed to be the same as the industry lead time (ILT). On the other hand, Buffer Management (BM) is a guiding tool for synchronous manufacturing. The time buffer is divided into three different zones (green, yellow, and red), each representing one-third of the total length. This is the execution part (in logistics category) and considered as an assessment tool for the production managers to monitor the system and intervene whenever the process is in danger (Schragenheim and Dettmer, 2000).

In this paper, an economic model for determining the length of the time buffer is developed based on Taguchi's loss function. The paper is organized as follows. The next section represents a review of literature pertaining to methods for determining the time buffer. Section 3 represents the steps taken in formulating the economic model, based on which the buffer size is derived. A numerical example is represented in Section 4 to illustrate the performance of the model. This is followed by an investigation of the model sensitivity to changes in its input parameters. Conclusions and final remarks are presented in Section 6.

2. Literature Review

The Time Buffer plays a significant role in DBR and SDBR implementations. It must be constantly monitored in order to avoid disruptions. Therefore, it is considered both an execution component and a warning alarm for managers. The buffer mechanisms in both DBR and SDBR share a common goal of minimizing constraint idleness resulting from upstream disruptions, thereby preventing throughput loss (Ye and Han, 2008). According to Schragenheim et al. (2009) "The buffer is the protective device used to ensure that a CCR is not starved and its capacity lost. It represents the amount of time in advance of need that work-in-process is scheduled to arrive at a particular control point."

Goldratt (1990) emphasized that the time buffer determination is not easily acquired and must be selected carefully because it may affect the net profit. According to Zhao and Hou (2014), "the smaller the time buffer, the more likely that it will not protect the CCR from some of the disruptions in the production process. But a too-large time buffer can cause a significant waste of resources by accumulating unnecessary inventory stock in the fort of the CCR."

Approaches for determining the length of time buffer appear to fall under one of two approaches; heuristic or quantitative.

2.1. Heuristic Approaches

Goldratt (1988) suggested that an initial buffer estimate can be obtained by taking one-half the current lead time. This estimated buffer size can be adjusted up or down according to the number of jobs that require expediting. Schragenheim and Ronen (1990) proposed that the buffer may be triple the constraint's average lead time. They pointed out that this value is drawn from their experience and contingent on realistic lead times.

Chase and Aquilano (1995) recommended using one-fourth of the total lead time of the system as an initial time buffer. Schragenheim (2006), suggested that the addition of 50% of the total lead time to the planned load is enough to meet the customer's expected delivery date, but not to deliver earlier than promised. Schragenheim et al. (2009) indicated that the buffer size in SDBR should represent a "liberal estimate" of the time segment between the release of the manufacturing components and the completed product's arrival to the shipping dock. They also indicated that most shipping buffers in make-to-order situations should not exceed the quoted lead time. However, Spencer (1991) recommended the use a trial and error approach in determining the buffer size.

2.2. Quantitative Approaches

Radovilsky (1998) developed an economical model that allows a trade-off between throughput and operating expenses. Following the notation of TOC, he estimated the net profit as the difference between throughput and operating expenses. He formulated the model assuming exponential service and Poisson arrivals for a single server with finite queue of no more than K units. Under the assumption that the arrival rate is equal to the service rate ($\rho = 1$), he developed a model that takes the general form of the economic order quantity formula. However, he did not address scenarios where arrival and service rates are not equal. Adetunji et al. (2011) extended the applicability of the same model (Radovilsky, 1998) to the cases where $\rho < 1$. They determined the buffer size required to achieve an optimal material flow in the system. The process is assumed to follow the $M/M/1/\infty$ queue and that the steady state queue length represents the optimal buffer size.

Tu and Li (1998) proposed a model for determining the time buffer required to protect the constrained resource. The model is focused on the operational flow of products and the relationship between the constrained resource and its feeders. The time buffer required to protect the constraint was expressed as a function of the mean time between failures and the mean time to repair all the feeder machines. Due to the complexity of the model, they proposed a 6-step algorithm for determining the maximum desirable buffer size at a stated level of confidence. They employed numerical examples and simulation experiments to validate their model. Similar work was presented by Ye and Han (2008). They reported that the more the feeder machines, the larger the buffer size needed.

Louw and Page (2004) used an open queuing network approach to estimate the length of the time buffer. The procedure starts with the demand and demand variability for each product. The routing for each product through the network is characterized by a production flow network that is determined by a computerized procedure that combines bill-of-material and routing data for each product as input by the user. Workstation data specifying the capacity of each resource as well as its mean time between failures and mean time to repair are also required. The averages and variances of flow time along the longest flow time are calculated and added to obtain an estimate of the total time buffer length as:

$$\textit{Buffer length} = \textit{Average flow time} + Z\,(\textit{Variance of flow time})^{1/2} - \textit{Sum of processing times}$$

Where; Z is the standard normal variable corresponding to the target service level. The procedure requires accurate estimation of 18 system parameters to provide an initial estimate of the time buffer length.

Also, Zhao and Hou (2014) utilized the queuing theory to study the conflict between unpredictable arrivals and finite-capacity resources. The resulting model requires knowledge of 6 queue parameters and 7 performance variables. It makes use of the mean and standard deviation, or the coefficient of variation of the inter-arrival and service time distributions as well as the number of servers. A spreadsheet was developed to help obtain approximate values. Approximations were compared with results from a Monte Carlo simulation at various levels of utilization. They indicated that the quality of approximation was not the same for all cases.

2.3. Discussion

The most important goal in managing a process is to minimize total cost while increasing customer satisfaction. One of the difficulties in achieving this goal is variability. Variability may negatively impact the performance of the system. To protect internal activities and assure smooth operations, adequate buffer needs to be allowed. Without appropriate buffering, the total costs will increase causing the total net profit to decrease. Under the Synchronous manufacturing philosophy, the physical inventory or WIP is eliminated by establishing a time buffer.

Based on the literature in the preceding section, the heuristic approaches for determining the size of the time buffer are simple and require knowledge of the total lead time only. However, they are supported by experience without quantitative justifications. The underlying assumption common to these heuristic approaches is that the lead time is very small compared to the industrial lead time (ILT). It is commonly assumed that there is always room to add slack (buffer) to the actual lead time. While this may be true for some markets, it is not the general case. Customer demand for short lead time has been a driver for competition over the last decade. Fewer and fewer customers are willing to wait for delivery if the product is readily available from the competitor. In addition, the idea of trying different buffer sizes to determine the appropriate size through trial-and-error during actual production would increase variability and cause unwarranted losses. Also, as was noted by Radovilsky (1998), if the initial buffer size was selected without preliminary evaluation, it may take a significant amount of time to decrease it to an appropriate level.

The quantitative approaches cited provide models that can be classified as economic (Radovilsky, 1998; Adetunji et al. 2011) or analytic (Tu and Li, 1998; Louw and Page, 2004; Ye and Han, 2008; Zhao and Hou, 2014). Economic models allow a trade-off between the system's throughput and the inventory carrying costs. Both have been expressed as functions of the buffer size under specific assumptions regarding the process model. This has restricted development efforts to single-product, single-server systems with Poisson arrivals and exponential service times. In addition, the models do not account for the effect of disruptions in upstream resources on arrivals at the CCR, as was noted by Louw and Page (2004).

In addition, the cited economic models do not account for any delay penalties that could be incurred when the quoted due dates are violated. These penalties could be substantial and cannot be ignored. In addition, the models do not account for losses resulting from sacrificing system capacity. It is implicitly assumed that such losses are much lower than the inventory carrying costs and that the excess capacity in non-constrained resources cannot be utilized in any other way.

On the other hand, the analytical approaches attempt to quantify the magnitude of the variability in the system (i.e., Murphy) to determine the size of the time buffer. This involves extensive modeling efforts leading to complex formulas more likely to repel practitioners than guide their applications. The model proposed by Tu and Li (1998) and extended by Ye and Han (2008) assumes stability of all stations. The time buffer is allocated based on the failure rates of the feeders preceding the CCR. Larger time buffers are allowed to absorb variations in the input. This additional time will enlarge the total lead time and contribute to extra costs. In fact, a large time buffer may cause the CCR to become idle. This is in sharp contrast with the principles of TOC, which attempt to exploit the CCR to assure its full utilization.

The procedure proposed by Louw and Page (2004) utilized Whitt's (1983) queuing network analyzer. They used queuing formulas to estimate flow time parameters within the entire network. This requires users to specify product demand data, bill-of-material, routing and workstation data including MMTF and MTTR for each workstation within the network. However, this procedure would work only for existing and stable processes with known parameters. Changes in demand, product mix or routing are expected to result in different estimates. Also, the procedure does not support applications of the SDBR, which allocates the time buffer at the end of the production line.

The quantitative model proposed by Zhao and Hou (2014), also utilized the queuing theory in an effort to study the conflict between unpredictable arrivals and finite-capacity resources. Their modeling procedures are based on the same assumptions made by Louw and Page (2004). They further assumed that customers are served on a first-in first-out basis. This is an unrealistic assumption, since the customers in a TOC managed system are served based on priorities set by the status of the time buffer. The time buffer is typically divided into three zones and based on the buffer status, priorities of service are established.

3. Model Formulation

The size of the time buffer has been addressed in several different ways as mentioned in Section 2. However, in this paper, we develop a new economic model for determining the time buffer by utilizing Taguchi's quadratic loss functions. In particular, we incorporate two functions to calculate the total losses associated with a given buffer size. The first (smaller is better) represents the relationship between the system capacity and the time buffer. The second function (larger is better) represents the relationship between the delay penalties and the size of the time buffer.

The models developed by Taguchi et al. (1989) are focused on estimating the losses resulting from deviations from the target values. The loss is proportional to the squared deviation from the target. This relationship is expressed in three different cases (target is best, smaller is better, and larger is better). The first case, assigns losses for deviations on both sides from the target value. The second case assigns losses for positive deviations from the target, whereas in the third case, losses are assigned to negative deviations only (Mitra, 1998).

3.1. Economic Modeling

This section represents the development of an economic model for determining the size of the time buffer. Two loss functions are developed and utilized to calculate total loss resulting from assigning a buffer of any size to protect the CCR. The first represents the relationship between the system capacity and the buffer size. This relationship accounts for losses resulting from synchronizing the production system to the CCR. The second loss function represents the relationship between the delay losses and the time buffer size. These two functions are incorporated into a total cost function in order to determine the economic buffer size. The notations used are represented in the following section.

3.2 Notations

A_1	Cost of idle capacity ($)
A_2	The incremental cost of expedited shipping ($)
ρ_1	Planned load on the CCR (time)
ρ_2	Amount of delay that can be tolerated by expedited shipping (time)
k_1	Proportionately constant for the capacity loss
k_2	Proportionately constant for the delay loss
$L_1(y)$	Capacity loss as a function of y ($)
$L_2(y)$	Delay loss as a function of y ($)
$TL(y)$	Total loss as function of y ($)
y	Time buffer (time)
α_{i}	Relative utilization for workstation (i)
λ_{i}	Capacity of a workstation (i)
λ_{CCR}	Capacity of the CCR workstation
t_{CCR}	Processing time on CCR (time/unit)
Q	Order quantity (units/order)
p	Proportion of nonconforming units under stable operation
Δ_{i}	Idle capacity at workstation (i)
v_{i}	Value added to the unit at workstation (i) ($/unit)
LT	The actual lead time for processing the order (time)

3.3 Capacity Loss Function

The concept of synchronous manufacturing was created to keep the CCR busy all the time to attain the maximum throughput. However, keeping all non-constraint resources under the control of the CCR with a time buffer in front of the CCR to prevent any disruption increases the likelihood of idle resources (upstream and downstream). In fact, the greater the buffer size, the more loss is incurred. This relationship can be represented by the smaller is better loss function, proposed by Taguchi in the form:

$$L_1(y) = k_1 y^2 \qquad (1)$$

Figure 1 illustrates the relationship between the time buffer (*y*) and the associated loss (*L*) per order as depicted in the *X*-axis and *Y*-axis respectively. The curve originates from the point of *y* = 0 at which L_1= 0. Over assigning the buffer will result in increased losses. These losses are attributed to the idle non-CCR resources. Therefore, the more the buffer length used for protection, the more the losses incurred. The constant K_1 is the proportionality constant representing the loss rate. Assuming that the loss associated with a target buffer of ρ_1 is A_1, then according to Equation 1;

$$A_1 = k_1 \rho_1^2 \qquad (2)$$

That is, the proportionality constant K_1 is given by:

$$k_1 = \frac{A_1}{\rho_1^2} \qquad (3)$$

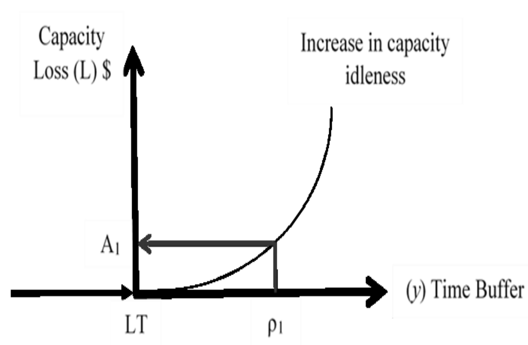


Figure 1: The capacity loss function

Practically, quoted lead time (QLT) is the sum of the time buffer (*y*) and the total lead time (LT), which is needed to process the order as depicted in Figure (2). As shown below, the QLT is the quote given to the customer regarding delivery date. It is important to note that QLT should be established at or below the industry lead time (ILT). Quotes exceeding the ILT are not typically accepted by customers and may result in excessive losses.

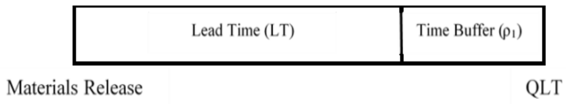


Figure 2: The lead time and time buffer

The value of ρ_1, can be obtained as a function of the order size Q, and other key performance characteristics of the CCR. Assuming that the unit processing time at the CCR is t_{CCR} and that the proportion of nonconforming units made under stable levels of operation is p ($p \geq 0$), then a possible value of ρ_1 can be obtained as:

$$\rho_1 = Q\,(1+p)\,t_{CCR} \qquad (4)$$

Basically, ρ_1 represents the planned load as defined by (Schragenheim & Dettmer, 2000). However, the value of ρ_1 above, considers the processing time of the CCR and the throughput of the system. All non-CCR resources contain residual capacities that have not been utilized as shown in Figure 3. The idle capacity Δ_i for each non-CCR resource can be determined as:

$$\Delta_i = (1-\alpha_i) \qquad (5)$$

Where α_i is the relative utilization for resource (i). This is computed by dividing the capacity of the CCR (λ_{CCR}) by the capacity of each non-CCR resource (λ_i) as:

$$\alpha_i = \frac{\lambda_{CCR}}{\lambda_i} \qquad (6)$$

As such, the loss associated with a buffer of size ρ_1 is given by:

$$A_1 = Q\,[\textstyle\sum_{i=1}^{m} \Delta_i\, v_i] \qquad (7)$$

Where, v_i is the value added by each non-CCR resource if the idle capacity was utilized. This value represents revenue that could have been generated by the firm if all the resources were completely utilized.

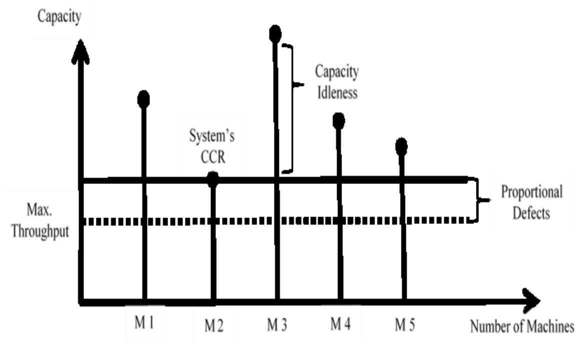


Figure 3: The idle capacity diagram

3.4 Delay Loss Function

On the other hand, the second function determines the relationship between the buffer size and the delay costs. Again, this relationship can be modeled by utilizing the larger is better loss function developed by Taguchi, which is given by:

$$L_2(y) = k_2 \frac{1}{y^2} \qquad (8)$$

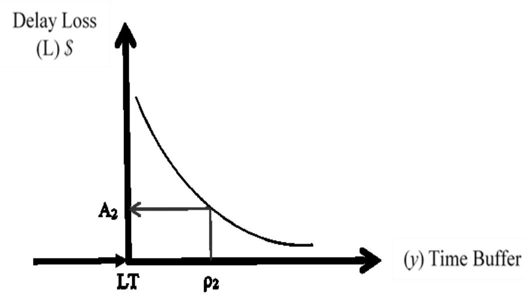


Figure 4: The bigger buffer is better

As depicted in Figure 4, the function represents infinite losses when no buffer (y=0) is allowed. The higher the buffer y assigned, the higher the chances of on-time delivery (and consequently the lower the loss). Applying this concept to the delay loss function requires the calculation of the proportionality constant K_2. Assuming that ρ_2 is the delay time that can be compensated for through expedited shipping at an additional (incremental) cost of A_2, then:

$$k_2 = A_2 \rho_2^2 \qquad (9)$$

Establishing an appropriate level of k_2 requires a close consideration of the cost schedule offered by various carriers. For example a delay of 2 days may be compensated for by changing the transportation mode at some additional cost.

3.5 Total Loss Function

Based on the above, the total loss associated with any time buffer *y* can be calculated as:

$$TL(y) = L_1(y) + L_2(y) \qquad (10)$$

Substituting for $L_1(y)$ and $L_2(y)$ from Equations (1) and (8), the economic buffer size appears at the intersection of the two functions as shown in Figure 5. At this point,

$$y_* = \left(\frac{A_2}{A_1}\right)^{0.25} (\rho_1 * \rho_2)^{0.5} \qquad (11)$$

And the minimum loss at y_* is given by:

$$TL(y_*) = \frac{A_1}{\rho_1^2} (y_*)^2 + \frac{A_2\rho_2^2}{(y_*)^2} \qquad (12)$$

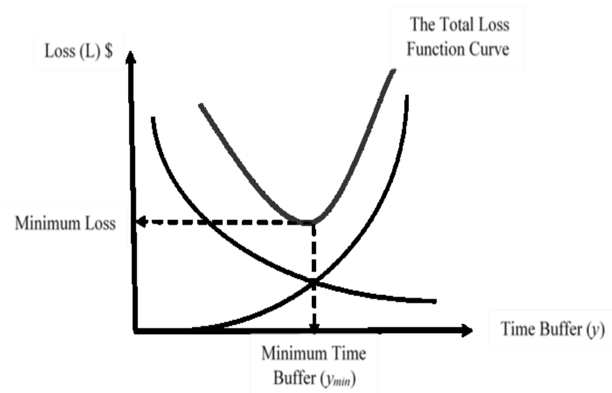


Figure 5: The Total Loss Function

4. Illustrative Application:

As an example, consider the application represented by Alghamdi et al (2015). This involves a computer assembly shop with five workstations in which M_2 was identified as the CCR when processing PC orders of specific configuration. The unit processing time on M_2, (t_{CCR}) is 1.5 hours. The fraction of nonconforming units made under stabled operation was estimated as 5%. Synchronizing all workstations to M_2 resulted in $\alpha_1 = 0.50$, $\alpha_3 = 0.38$, $\alpha_4 = 0.63$ and $\alpha_5 = 0.77$. Each workstation contributes an equal amount to the added value of \$20 per unit. For an order of size of *Q*= 100 units and according to Equation (7), (A_1) amounts to \$3440. By using Formula (4), ρ_1 is calculated as 157.5 hours. On the other hand, orders that suffer a delay of ρ_2= 24 hours can be expedited at an additional cost of A_2= \$100.

Using Equations (11) and (12), the buffer size is determined as 25.4 hours at which a minimum loss of \$178.7 is achieved.

This buffer should be added to the estimated lead time in establishing the QLT. In applications of the SDBR, this represents the shipping buffer. Whereas, in the traditional DBR the total buffer from Equation (11) should be divided equally into three components representing buffers for the CCR, assembly and shipping.

5. Model Performance

To evaluate the performance of the model, a factorial experiment is utilized to investigate the effect of changes in the input parameters on the buffer size. In setting up the experiment, the application in the previous section was utilized. Six independent parameters were identified as shown in Table (1). The statistical design used is an un-replicated 2^6 factorial with 64 runs. The values of θ are used to represent values of ($\sum \Delta i\ Vi$) used in Equation (7). Each factor was assigned two levels at ± 10% the nominal level used in the example. Over all, the changes resulted in an average buffer of 25.3 with standard deviation of 2.12 hours, and average loss of \$180.8 with standard deviation of \$30.4. From a practical standpoint, the model appears robust to errors in estimating values of its parameters (within ±10%). However, statistical analysis of the results revealed that the model is sensitive to changes in all six parameters.

Table 1: Model Parameters and Their Levels

Model Parameter	Level	
	Low	High
A_2	90	110
ρ_2	21.6	26.4
tCCR	81	99
Q	90	110
p	0.045	0.055
θ	30.96	37.84

The half-normal probability plot shown in Figure 6 indicates that both processing time at the CCR (tCCR) and the amount of delay that can be tolerated by expedited shipping (ρ_2) have the highest positive effects on the average buffer size. Both contributed 72% of the total variability in the average buffer. The analysis of variance (ANOVA) has confirmed these results. This suggests that practitioners need to pay special attention when estimating values of these parameters in future applications of the proposed model.

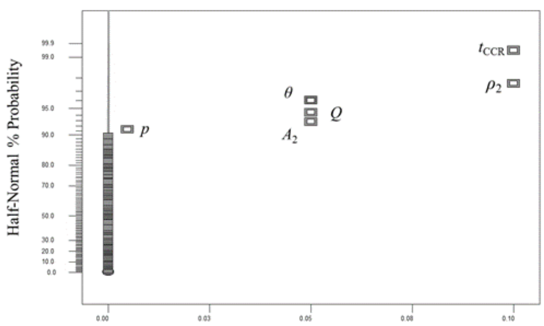


Figure 6: Half-Normal Plot of Absolute Effects

6. Conclusions

In Synchronous manufacturing or theory of constraints, the essential goal is to maximize profits by utilizing the CCR to the highest extent possible. This requires an allocation of a time buffer to protect the CCR from starvation. In this paper, an economic model was developed to help determine the buffer size. The model allows a tradeoff between capacity and delay losses, both of which are minimized by using Equation (11). The model requires estimates of six parameters only and allows the determination of a buffer size with the minimum associated loss. As such, it is in line with the simplicity of the DBR approach, which appeals to its users and reduces the need for heuristic methods. Also, no assumptions were made regarding the relative contribution of the CCR to the total value of the product. The proposed model accounts for the price (loss) of synchronizing the entire production system to the CCR. In contrast to the models cited in Section 2, users of the proposed model do not need to assume that the process lead time can always be extended. The new model does not require estimates of the losses associated with violations of the QLT. These are external losses that can derive customer dissatisfaction and are typically hard to estimate. In addition, the model does not assume perfect processing and accounts for the proportion of defective units that may be produced.

These authors are considering applications where processes are characterized by random variables with specified probability distributions. Their research efforts are aimed at utilizing probabilistic estimates of the process parameters and accounting for the losses associated with their variances.

References

Adetunji, O., Adendorff, K., & Yadavalli, V. S. S. (2011). Dynamic Buffering of a Capacity Constrained Resource via the Theory of Constraints. Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management, Kuala Lumpur, Malaysia, January 22-24, 2011.

Alghamdi, F, Weheba, G, and Abdulaziz A. (2015). Application of Drum-Buffer-Rope (DBR) in a Computer Assembly Plant. The Journal of Management and Engineering Integration 8(1), 1-8.

Chase, R. and Aquilano, N., Production and Operations Management: Production and Services, 1995 (Irwin: Homewood, IL).

Cox, J. F., & Schleier, J. G. (Eds.). (2010). Theory of constraints handbook. New York: McGraw-Hill.

Dettmer, H. W. (2000). Constraint Management. The Complete Guide to the CQM, 1-27. Retrieved from(ttp://www.goalsys.com/books/documents/ConstraintManagement.pdf)

Frazier, G. V., and Reyes, P. M. (2000). Applying synchronous manufacturing concepts to improve production performance in high-tech manufacturing. Production and inventory management journal, 41(3), 60.

Taguchi, G., Elsayed, E. A., & Hsiang, T. C. (1989). Quality engineering in production systems. McGraw-Hill College.

Goldratt, E. M. (1988). Computerized shop floor scheduling. International Journal of Production Research, 26(3), 443-455.

Goldratt, E. M. and Cox, J. (1992). The Goal: A Process of Ongoing Improvement, North River Press, Croton-on-Hudson, New York.

Goldratt, E.M. (1990), What is This Thing Called Theory of Constraints and How Should it be Implemented?, North River Press, New York, NY

Groop, J. 2012. Theory of constraints in field service: factors limiting productivity in home care operations. Dissertation, Aalto University, Espoo, Finland.

Louw, L., & Page, D. C. (2004). Queuing network analysis approach for estimating the sizes of the time buffers in Theory of Constraints-controlled production systems. International Journal of Production Research, 42(6), 1207-1226.

Mitra, A. (1998). Fundamentals of quality control and improvement, 2nd edition Prentice-Hall.

Peltokorpia, A., Nis_enb, H., Groopb, J. Reinikainenc, T., Bengsc, A., and Pirttimaa, M. (2016). Applying the Theory of Constraints to Improve Throughput in a Forensic DNA Laboratory. Forensic Science Policy & Management, 7 (1–2), 37–49.

Radovilsky, Z. D. (1998). A quantitative approach to estimate the size of the time buffer in the theory of constraints. International Journal of Production Economics, 55(2), 113-119.

Rahman, S. U. (1998). Theory of constraints: a review of the philosophy and its applications. International Journal of Operations & Production Management, 18(4), 336-355.

Robbins W. A. (2011). Process Improvement in the Public Sector: A Case for the Theory of Constraints. Journal of Government Financial Management, 40-46.

Sadat S., Carter M.W. and Golden B. (2013). Theory of constraints for publicly funded health systems. Health Care Management Science, 62-74.

Schragenheim, E. (2006). Using SDBR in rapid response projects. The Goldratt group. http://web.ftfsys.com/zl/Using%20SDBR%20in%20Rapid%20Response%20Project.pdf

Schragenheim, E., & Dettmer, H. W. (2000). Simplified Drum-Buffer-Rope A Whole System Approach to High Velocity Manufacturing. Goal Systems International.

Schragenheim, E., & Ronen, B. (1990). Drum-buffer-rope shop floor control. Production and Inventory Management Journal, 31(3), 18-22.

Schragenheim, E., Dettmer, H.W. & Patterson, J.W. (2009). Supply chain management at warp speed: Integrating the system from end to end 1st ed., Auerbach Publications.

Siha, S. (1999). A classified model for applying the theory of constraints to service organizations. Managing Service Quality, 9(4), 255-264.

Sivasubramanian, R., Selladurai, V., & Gunasekaran, A. (2003). Utilization of bottleneck resources for profitability through a synchronized operation of marketing and manufacturing. Integrated manufacturing systems, 14(3), 238-246.

Spencer, M. S. (1991). Using 'The Goal' in an MRP system. Production and Inventory Management Journal, 32(4), 22.

Spencer, M. S., & Cox, J. F. (1995). Optimum production technology (OPT) and the theory of constraints (TOC): analysis and genealogy. Advances in Engineering & Technology, 3(1), 334-344.

Watson, K. J., Blackstone, J. H., & Gardiner, S. C. (2007). The evolution of a management philosophy: The theory of constraints.

Journal of Operations Management, 25(2), 387-402.

Ye, T., & Han, W. (2008). Determination of buffer sizes for drum–buffer–rope (DBR)-controlled production systems. International Journal of Production Research, 46(10), 2827-2844.

Zhao, X., & Hou, J. (2014). Analyzing the time buffer in the Theory of Constraints based lean operations. Journal of Management Analytics, 1(3), 185-199 THE International Journal of Production Research, 33(6), 1495-1504.

Tu, Y. M., & Li, R. K. (1998). Constraint time buffer determination model. International Journal of Production Research, 36(4), 1091-1103.

Tulasi, C. H., & Rao, A. R. (2012). Review on theory of constraints. International Journal of

.

Quality in Higher Education: Perceptions of Top Administrators

Faisal Alotaibi[1], Gamal Weheba[1], and Serkan Toy[2]
[1]*Department of Industrial and Manufacturing Engineering, Wichita State University*
Wichita, KS 67260
[2]*Department of Anesthesiology, School of Medicine, University of Kansas*
Wichita, KS
67214
gamal.weheba@wichita.edu

Abstract

Quality in higher education is driven by numerous factors. One of these factors is the top administrators' perceptions of quality. In their communication to the public, administrators' state initiatives and stress dimensions that they believe would increase stakeholders' satisfaction and derive institution's reputation. This study examines the administrators' views on quality in the top 100 ranked universities in theUS. A content analysis of letters from university presidents revealed the most important dimensions perceived to affect quality in higher education. The steps involved in identifying these dimensions and determining their significance are represented. It is hoped that these research findings help provide a clear understanding of the dimensions of quality, and direct improvement efforts in such a critical service sector. Also the paper concludes with practical implications and proposes directions for further research.

1. Introduction

Studies on quality in higher education have been diverse. An initial title search of the library database utilizing the words "service quality" along with "higher education," "university," or "college" identified 4,059 peer-reviewed journal publications between 2005 and 2015. Research in this sector presents higher education as a service. The intangibility of the outcome, heterogeneity of the process, and inseparability of production and consumption have been cited in support of such a classification. However, some researchers have viewed higher education as an atypical service industry because of the aspects incorporated and the multiplicity of stakeholders. Furthermore, unlike many service sectors, higher learning institutions have stringent academic and sometimes personal entry requirements that define the population to which the service is rendered; this creates a specific customer base (Rowley, 1997).

In this paper, Computer-Aided Text Analysis (CATA) and multivariate analysis techniques have been utilized to determine the dimensions of quality as perceived by university administrators. These perceptions are likely to affect operations and directions for higher education institutions within the U.S. Also of interest is the relative weight that the administrators assign to each dimension in their communication with the public. The following section presents a review of the literature pertaining to quality in service and higher education. Section 3 offers a description of the methodology used and results obtained. Final conclusions and ongoing research efforts are highlighted in Section 4.

2. Literature Review

Garvin (1984) noted that definitions of quality may fall into one of five categories: transcendental, product-based, user-based, manufacturing-based, and value-based. He developed eight dimensions which can be used to evaluate quality. These are performance, features, reliability, conformance, durability, serviceability, aesthetics, and perceived quality. He indicated that these dimensions can be used to evaluate quality of products and services. Grönroos (1984) defined service quality as having two main variables: consumers' expectations in term of outcome, and perceptions of the results.

Parasuraman et al. (1985) have provided a list of ten dimensions to be considered in service quality. Their study included samples of customers of retail banking, credit cards, securities brokerage, and repair shops. They defined service quality as the ability of an organization to meet or exceed customer expectations. In other words, quality can be measured by the difference between customers' initial expectations and their perceptions after receiving the service.

By using factor analysis, Parasuraman et al. (1988) reduced the proposed dimensions into five: reliability, responsiveness, tangibles, assurance, and empathy. Three dimensions remained the same from the original ten (reliability, responsiveness, and tangibles), and the remaining seven dimensions from the original were consolidated into two (assurance and empathy). Based on these five dimensions, a 22-item survey instrument for measuring service quality was developed. This tool, called the SERVQUAL, is utilized to measure the gap between customer's perceptions of the service and their expectations.

Owlia and Aspinwall (1996) represented a comparison between the original dimensions proposed by Parasuraman et al. (1985) and five other service quality models. They proposed a framework for defining quality in higher education based on the six dimensions. These included tangibles, competence, attitude, content, delivery and reliability.

Cheng and Tam (1997) suggested that quality in higher learning institutions can be defined by a set of components in the input, process, and output that satisfy explicit and implicit expectations. In an effort to standardize quality assessments, they proposed a seven-component model. These are goal and specification, resources, processes, satisfaction, legitimacy, absence of problems, and organizational learning.

LeBlanc and Nguyen (1997) examined service quality in higher education and proposed seven dimensions: reputation, administrative personnel, faculty, curriculum, responsiveness, physical evidence, and access to facilities.

Abdullah (2006) noted the need for more encompassing tools. He pointed out that quality entails more than academic competency and includes aspects of the total service environment as experienced by the students. He also noted the positive relationship between quality standards and such aspects as increased profitability, customer satisfaction, customer loyalty, customer retention, customer attraction, and positive word of mouth. Based on this, he proposed a scale for assessing quality known as the Higher Education Performance (HEdPERF) scale.

Mahapatra and Khan (2007) proposed the EduQUAL instrument for measuring the critical factors of students' perceived service quality by using learning outcomes, responsiveness, personality development and academics.

Faganel (2010) observed that efforts to

deliver quality in higher education are often focused on one specific area, such as teaching, and then the findings are projected to represent the overall quality of the institution, ignoring aspects of administrative services. He suggested that improvements in clarity, accuracy, and reliability of services offered by instructors are fundamental to the overall quality.

Annamdevula and Bellamkonda (2012) developed a measurement scale to evaluate the quality of service in higher education, known as the Higher Education Service Quality (HiEdQUAL) scale. Their study considered only students as the main customers in the sector. This scale has 27 items containing five dimensions including teaching and course content, administrative services, academic facilities, campus infrastructure, and support services.

In addition, higher education accreditation in the U.S. is achieved through a peer-review process by an accredited organization. The process is set up to ensure that quality is maintained and sustained through all registered higher learning institutions. This is achieved through rigorous checks on curriculums, equipment, and services, both in registered learning institutions and in institutions seeking registration (Eaton, 2006). Therefore, the process of accreditation in the U.S. serves as a quality assurance system.

3. Methodology

This section describes the methodology and research techniques utilized to identify dimensions used to describe quality from the viewpoint of top administrators in higher education. The research involved content analysis of a sample of letters from presidents of top ranked universities in the U.S. As was pointed out by Berelson (1952), content analysis is utilized in research for systematic, objective, and quantitative description of the manifest communication content. Computer-Aided Text Analysis refers to the use of computer applications to analyze wording used in passages and surveys. It is useful for analyzing and interpreting text. This type of analysis allows researchers to take advantage of automated processes in making guided conclusions based on textual contents of publications. It also saves time and improves accuracy. These procedures can be applied to a large set of data and still maintain high reliability (Duriau et al., 2007). The following subsections describe the steps taken as was recommended by Short et al., (2009) to ensure accuracy.

3.1 Initial Constructs

The first step in this research utilized the results of the literature review represented in Section 2. Ten dimensions of service quality from Parasuraman et al. (1985), and three dimensions from Garvin (1984) have been selected as initial constructs. These thirteen dimensions (constructs) are listed in Table 1.

In defining each construct, the term stakeholder was used rather than customer. Definitions of the constructs proposed by Garvin (1984), including performance, features, and conformance were modified by using the term institution in place of corporation and/or product.

3.2 Coding Scheme

In the second step of this research, a list of words (codes) for each construct was created, based on the summary of applications of three recipients of the Malcolm Baldrige National Quality Award in higher education: the University of Wisconsin-Stout in 2001, Kenneth W. Monfort College of Business in 2004, and Richland College in 2005. These are available on line from http://www.nist.gov/baldrige. A list of the most frequent words used by MBNQA recipients was generated, and sorted by two higher education experts.

Table 1. Initial Constructs and Their Definitions

Construct	Definition
Reliability	Consistency of performance and dependability; means performing the service right the first time and that the institution honors its promises.
Responsiveness	Willingness and readiness of faculty and staff to provide service.
Competence	Possession of required skills and knowledge to perform the service.
Access	Approachability and ease of contact.
Courtesy	Politeness, respect, consideration, and friendliness of contact personal.
Communication	Keeping stakeholders informed and listening to them.
Credibility	Trustworthiness, believability, and honesty.
Security	Freedom from danger, risk, or doubt.
Understanding	Making an effort to understand stakeholders' needs.
Tangibles	Physical evidence of service.
Performance	Primary operating functions of the institution.
Conformance	Extent to which the institution meets pre-established standards (both internal and external).
Features	Supplemental characteristics offered by the institution.

A special survey was designed and used to validate these codes. The survey was completed by five members of the faculty as content experts. Based on the responses obtained, a content validity ratio (CVR) was calculated as per Lawshe (1975). Codes with CVR below 0.99 were excluded from the list. In order to avoid bias among the constructs, the total number of codes for each construct was unified to five codes. The final list of codes used to perform the CATA is shown in Appendix A.

3.3 Textual Analysis

In determining administrator's perceptions of quality in higher education, the top 100 universities were identified from the U.S. News and World Report's rankings (2015). The list of universities included is shown in Appendix B. This sample size was determined under the assumption that all constructs are equally likely to be used and that a margin of error of 5% is acceptable with a 95% confidence level. Based on this list, presidents' letters were downloaded from each university's website. The full text of these letters were analyzed, using the NVivo software, to determine the frequency of occurrence of each construct and its codes. The results were used to construct a 100 × 13 contingency table. Rows represented universities and columns represented the constructs. A chi- square test was applied to test for independence. The test supported the assumption of independence with a P-value < 0.0001. As such, principal components analysis was applied using the STATGRAPHICS software and the results indicated that the top four components contribute 72.7% of the total variability as shown in Table 2.

Table 2. Principal Components

Comp. Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	5.61316	43.178	43.178
2	1.54205	11.862	55.040
3	1.28315	9.870	64.910
4	1.01554	7.812	72.722
5	0.704163	5.417	78.139
6	0.626856	4.822	82.961
7	0.503094	3.870	86.831
8	0.409285	3.148	89.979
9	0.338769	2.606	92.585
10	0.305424	2.349	94.935
11	0.271826	2.091	97.026
12	0.228274	1.756	98.782
13	0.158405	1.218	100.000

The first principal component contributes 43.1% to the variance. An examination of the components' weights shown in Table 3 indicates that access, communication, understanding, and performance are the dimensions with significant contributions to the first principal component. These are functional aspects that relate to interactions between the institution and its stakeholders and are often perceived subjectively. According to Nelson (1974), these are experience properties that can only be assessed after or during service. This suggests a dimension of "empathy" because they mostly deal with "support services" as defined by Annamdevula and Bellamkonda (2012). In this context, communication appears to correlate with understanding and access.

The second principal component, shown to contribute 11.8% to the variance, suggests a dimension of "attitude" since it includes behavioral aspects such as courtesy, responsiveness and conformance. These are also functional aspects, often perceived subjectively, and can only be evaluated during or after interaction. This is supported by Owlia and Aspinwall (1996).

Security and reliability are included in the third principal component, which accounts for almost 10% of the variability. These can be considered aspects of "safety" as they relate to campus environment and the ability to protect individuals, properties and data over time.

The fourth component includes aspects of competence, credibility, features and tangibles, accounting for 7.8% of variability. These are mostly image-related aspects that correlate with the "reputation" of the institution as perceived by the stakeholders. Such dimension was proposed by LeBlanc and Nguyen (1997). This dimension includes search aspects. Search aspects involve technical properties that can be evaluated without interaction such as tangible and features, as well as competence and credibility. They are typically perceived subjectively form the university website and other publications. Table 4 lists dimensions that made significant contributions to each component.

4. Conclusions

This research has indicated that all thirteen constructs have been utilized by administrators in their communication with the public. It also suggests some meta-dimensions of quality in higher education including: empathy, attitude, safety, and reputation. These results indicated that perceptions of top administrators appear to be dominated by experience aspects.

Table 3. Components' Weights

Constructs	Component 1	Component 2	Component 3	Component 4
Reliability	0.260864	-0.0268606	**-0.266317**	-0.0684558
Responsiveness	0.229578	**0.507153**	0.275682	-0.026998
Competence	0.219368	-0.0912024	0.390875	**0.513426**
Access	**0.326378**	-0.0687589	-0.166728	-0.00452369
Courtesy	0.173385	**0.596864**	0.194777	0.073251
Communication	**0.344548**	0.249459	-0.102271	-0.153678
Credibility	0.208005	-0.164936	0.437692	**-0.517424**
Security	0.239285	0.201616	**-0.452053**	-0.23706
Understanding	**0.332557**	-0.169381	0.247567	-0.145742
Tangibles	0.276368	-0.211273	-0.257226	**0.279873**
Performance	**0.342797**	0.00624929	-0.262416	0.173204
Conformance	0.261955	**-0.360795**	0.0820319	-0.336808
Features	0.319682	-0.197891	0.158506	**0.36748**

Table 4. Principal Components and Constructs

PC 1	PC 2	PC 3	PC 4
Communication	Courtesy	Security	Credibility
Performance	Responsiveness	Reliability	Competence
Understanding	Conformance		Features
Access			Tangibles

Empathy, attitude, and safety (the top three principal components) are dominated by functional aspects that relate to interactions with stakeholders. Performance has been used within the context of efficiency (utilization of funds and/or potentials for graduates). Such findings may motivate other administrators to follow suit and establish appropriate measures, goals and deploy plans allowing their institutions to score high on these dimensions. After all, successful quality improvements have always been achieved with support from top administrators.

It is important to note here that the dimensions identified in this research describe perceptions of quality based on the selected sample. It is expected that different samples would provide different results, just as much as different stakeholders would have different perceptions about quality. Time is also an important factor in defining aspects of quality. This research provided a static picture of current perceptions of top administrators. As expectations change over time, perceptions of quality are expected to follow. Repeated research efforts over appropriate periods of time would help detect changes and trends in the perceptions of quality in higher education. Similar research using online reviews and job announcements as units of analysis would also be beneficial in determining perceptions of students and employers. Such efforts would be facilitated by the growing number of social media users and the availability of data mining techniques.

References

Abdullah, F. (2006). "The development of HEdPERF: A new measuring instrument of service quality for the higher education sector." International Journal of Consumer Studies, 30 (6), 569–581.

Annamdevula, S., & Bellamkonda, R.S. (2012). "Development of HiEdQUAL for measuring service quality in Indian higher education sector." International Journal of Innovation, Management and Technology, 3 (4), 412–416.

Berelson, B. (1952). Content analysis in communication research. Glencoe, IL: The Free Press.

Cheng, Y. C., & Tam, W. M. (1997). "Multi-models of quality in education." Quality Assurance in Education, 5 (1), 22–31.

Duriau, V. J., Reger, R. K., & Pfarrer, M. D. (2007). "A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements." Organizational Research Methods, 10 (1), 5–34.

Eaton, J. S. (2006). "An overview of U.S. accreditation." Washington, DC: Council for Higher Education Accreditation.

Faganel, A. (2010). "Quality perception gap inside the higher education institution." International Journal of Academic Research, 2 (1), 213–215.

Garvin, D. A. (1984) "What does "product quality" really mean?" Sloan Management Review, 25–43

Grönroos, C. (1984). "A service quality model and its marketing implications." European Journal of Marketing, 18 (4), 36–44.

Lawshe, C. H. (1975). "A quantitative approach to content validity." Personnel Psychology, 28 (4), 563–575.

LeBlanc, G., & Nguyen, N. (1997). "Searching for excellence in business education: An exploratory study of customer impressions of service quality." International Journal of Educational Management, 11 (2), 72–79.

Mahapatra, S. S., & Khan, M. S. (2007). "Assessment of quality in technical education: An exploratory study." Journal of Services Research, 7 (1), 81–100.

Nelson, P. (1974). "Advertising as information." Journal of Political Economy, 82 (4), 729–754.

NVivo Qualitative Data Analysis Software (2014). QSR International Pty Ltd. Version 10.

Owlia, M. S., & Aspinwall, E. M. (1996). "A framework for the dimensions of quality in higher education." Quality Assurance in Education, 4 (2), 12–20.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). "A conceptual model of service quality and its implications for future research." The Journal of Marketing, 49 (4), 41–50.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). "SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality." Journal of Retailing, 64, 12–40.

Rowley, J. (1997). "Beyond service quality dimensions in higher education and towards a service contract." Quality Assurance in Education, 5 (1), 7–14.

Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. C. (2009). "Construct validation using computer-aided text analysis (CATA): An illustration using entrepreneurial orientation." Organizational Research Methods, 32, 320–347.

Statgraphics® Centurion (2009). Statgraphics® Centurion XVI user manual. Warrenton, VA: StatPoint Technologies, Inc.

U.S. News & World Report Education (2016). National universities rankings. URL: http://colleges.usnews.rankingsandreviews.com/best-colleges/rankings/national-universities [cited January 16, 2016].

Appendix A.
LIST OF CODES

Construct	Code
Reliability	Accuracy Completeness Confidence p Consistency Stability
Responsiveness	Flexibility Diversity Readiness Willingness Preparedness
Competence	Capability Experience Knowledge Skill Qualification
Access	Advising Affordability Approachability Availability Capacity
Courtesy	Accommodating Consideration Friendliness Politeness Respect
Communication	Contact Inform Interact Listen Participate
Credibility	Believability Ethical Honesty Integrity Trustworthiness

Construct	Code
Security	Assurance Confidentiality Protection Safeguard Safety
Understanding	Accept Assist Appreciate Cooperate Recognize
Performance	Developing Evaluating Improving Measuring Training
Conformance	Accredit Achieve Review Satisfy Update
Features	Curricula Degree Offerings Opportunities Program
Tangibles (Facilities)	Centers Classrooms Laboratories Libraries Equipment

APPENDIX B
List of top-ranked 100 Universities in the U.S.
Source: U.S. NEWS AND WORLD REPORT EDUCATION (2016)

Rank	University	Rank	University
1	Princeton University	25	University of Southern California
2	Harvard University	26	University of Virginia
3	Yale University	27	Tufts University
4	Columbia University	28	Wake Forest University
5	Stanford University	29	University of Michigan, Ann Arbor
6	University of Chicago	30	Boston College
7	Massachusetts Institute of Technology	31	University of North Carolina, Chapel Hill
8	Duke University	32	New York University
9	University of Pennsylvania	33	University of Rochester
10	California Institute of Technology	34	Brandeis University
11	Johns Hopkins University	35	College of William and Mary
12	Dartmouth College	36	Georgia Institute of Technology
13	Northwestern University	37	Case Western Reserve University
14	Brown University	38	University of California, Santa Barbara
15	Cornell University	39	University of California, Irvine
16	Vanderbilt University	40	University of California, San Diego
17	Washington University in St. Louis	41	Boston University
18	Rice University	42	Rensselaer Polytechnic Institute
19	University of Notre Dame	43	Tulane University
20	University of California, Berkeley	44	University of California, Davis
21	Emory University	45	University of Illinois, Urbana-Champaign
22	Georgetown University	46	University of Wisconsin, Madison
23	Carnegie Mellon University	47	Lehigh University
24	University of California, Los Angeles	48	Northeastern University

APPENDIX B (continued)

Rank	University	Rank	University
49	Pennsylvania State University	75	Clark University
50	University of Florida	76	Colorado School of Mines
51	University of Miami	77	Indiana University, Bloomington
52	Ohio State University, Columbus	78	Michigan State University
53	Pepperdine University	79	Stevens Institute of Technology
54	University of Texas, Austin	80	University of Delaware
55	University of Washington	81	University of Massachusetts, Amherst
56	Yeshiva University	82	Miami University, Oxford
57	George Washington University	83	Texas Christian University
58	University of Connecticut	84	University of California, Santa Cruz
59	University of Maryland, College Park	85	University of Iowa
60	Worcester Polytechnic Institute	86	Marquette University
61	Clemson University	87	University of Denver
62	Purdue University, West Lafayette	88	University of Tulsa
63	Southern Methodist University	89	Binghamton University, SUNY
64	Syracuse University	90	North Carolina State University, Raleigh
65	University of Georgia	91	Stony Brook University, SUNY
66	Brigham Young University, Provo	92	SUNY College of Science and Forestry
67	Fordham University	93	University of Colorado, Boulder
68	University of Pittsburgh	94	University of San Diego
69	University of Minnesota, Twin Cities	95	University of Vermont
70	Texas A&M University, College Station	96	Florida State University
71	Virginia Tech	97	Saint Louis University
72	American University	98	University of Alabama
73	Baylor University	99	Drexel University
74	Rutgers, State University of New Jersey	100	Loyola University Chicago

A Framework for Discovering Customer Requirements Based on Multidimensional Scaling

Muhammad Attar and Gamal Weheba
Industrial & Manufacturing Engineering Department, Wichita State University
Wichita, KS 67260
gamal.weheba@wichita.edu

Abstract

Multidimensional Scaling (MDS) and Cluster Analysis have great value in qualitative research. They allow for effective integration between team processes and statistical techniques. MDS enables researchers to cluster customer comments and merge data structures, in aims of tracking themes. It is one of the most useful techniques for analyzing customer feedback to determine their requirements. This paper represents an MDS-based framework for discovering new customer requirements. An example is used to illustrate the steps involved and guide future applications.

1. Introduction

There are several methods for determining customers' perceptions of the products offered or services provided as indicated by Juran and Godfrey (1999). These include direct face-to-face interviews, focus groups, and surveys. Customer feedback is key input for improving existing designs or offering new products or services. Valuable information about customer opinion can be elicited in an unstructured fashion (Erickson &Kaplan, 2000). Numerous studies have demonstrated the benefits of analyzing customers' comments and exploring their perceptions. Several concepts and requirements can be captured from their responses and comments (Miles & Huberman, 1994; Pothas, Andries & DeWet, 2001). According to Sproull (1988) qualitative text analysis can be used to determine dimensions that derive customer expectations.

The use of concept mapping (CM) to identify customer requirements has great value in qualitative research. Analyzing customer requirements using multidimensional scaling (MDS) is helpful in organizing customers' expectations and grouping them in an objective way. Indeed, when such information is accurately translated into technical characteristics by designers, quality products and services are more likely to be produced.

Unfortunately, research on qualitative data analysis of customer requirements has not been given much attention in recent years. Limited attempts have been made to develop a framework for discovering new requirements. This paper represents a new MDS-based framework that can be used to discover new requirements and emerging expectations.

The following section provides a review of the literature related to CM and MDS. Section 3 represents a description of the proposed framework. An illustrative application with a detailed procedure follows in Section 4.

2. Literature Review

Concept Mapping (CM) is a qualitative data analysis technique, which can be used to determine common themes in textual data. CM was first introduced by Jackson and Trochim (2002) as a qualitative method that includes a number of steps for collecting, analyzing, classifying, and visualizing data structures. The main feature of CM is that a group of sorters from the target population are involved in coding the data. The results represent the final units of analysis that are analyzed using MDS and cluster analysis. This approach was preferred by researchers over critical incidents and a semantic differential method due to the mathematical properties of the analysis used (Coxon, 1982, Kruskal and and Wish, 1978). MDS is a statistical technique used to analyze and visualize the data in multidimensional space. It allows for comprehension evaluation, dimensions identification, and interpretation of perceptions. According to Carroll and Wish (1976) MDS helps to project all data in multiple dimensions, provides graphical representations, and facilitates understanding of customer perceptions.

Kruskal and Wish (1978) demonstrated some of the advantages of using MDS. They pointed out that MDS can be utilized to analyze both metric and nonmetric data. Metric data is quantitative data such as interval or ratio data whereas nonmetric data is qualitative data such as a number of proximity (similarity or dissimilarity). Moreover, MDS maps provide meaning to patterns or regions with related variables.

According to Malhotra and Birks (2007), MDS involves six steps: formulate the problem, obtain input data, choose an MDS model, set number of dimensions, label the dimensions and interpret the configuration, and evaluate reliability and validity.

The main two difficulties of MDS applications are choosing an appropriate number of dimensions to represent the data, and interpreting the resulting structures. Choosing a large number of dimensions results in over-fitting the data. In contrast, choosing a small number of dimensions results in an inaccurate interpretation of data structure (Shinkareva,Wang & Wedell , 2013). A goodness of fit with the stress, and the coefficient of determination (R-squared) values are useful in obtaining an appropriate number of dimensions, measuring the adequacy of the MDS solutions, and obtaining information about how well the model fits the data. The smaller the stress value, the better the fit (Kruskal and Wish, 1978). A scree plot or scree test is commonly used where stress values are plotted against the number of dimensions. A clear elbow indicates an appropriate number of dimensions that need to be considered.

Malhotra and Birks (2007) indicated that the number of dimensions may be determined based on prior knowledge from past research, interpretation of the spatial map in two or three dimensions, elbow criterion that shows clear change in a certain point, ease of use of two dimensional maps, or by using statistical approaches. The interpretation of the MDS map is conducted through cluster analysis method in order to identify the structures of data within the map. These predictions do not change the MDS solution. However, confirmatory studies must be used to validate these interpretations. Borg and Groenen (2005), provide a useful discussion on confirmatory studies.

Hayes (1997) proposed a simple technique for establishing the quality (comprehensiveness) of the content of critical incidents. He proposed removing approximately 10 % of the comments from the initial list before the analysis. After the analysis is completed with 90% of the data, if the remaining 10% can be placed into the identified requirements, then the resulting list of requirements is comprehensive and includes all possible requirements. Otherwise the interpretation is deficient and more customers must be interviewed.

3. Proposed Framework

The proposed framework is shown schematically in Figure 1. It starts by collecting customer comments through direct interviews, focus groups, comment cards or surveys with open-ended questions, and identifying units of analysis. Long statements should be divided into simple units (comments) and ambiguous comments deleted. Resulting units are classified into current or emerging requirements. All emerging requirements (N) are compiled and saved in a document for further analysis. Sorters are invited to group similar units together. This involves grouping similar units in piles by each sorter. The results are scored on a 0, 1 frequency matrix. Cells in each N x N matrix contain 1 if the associated row and column were grouped together by the sorter in the same pile. Otherwise, a zero is assigned to the cell. This step results in the construction of one diagonal matrix for each participating sorter. These matrices are then combined by adding scores in each cell across all sorters resulting in an N x N similarity matrix. As such row (column) totals represent the level of agreement among sorters. The higher the total, the higher the level of agreement and vice versa.

Unlike the conventional approach, the matrix is reduced before performing MDS. A simple reduction technique calls for moving rows and columns with duplicated totals to another matrix. This results in constructing two different matrices, namely a reduced matrix and a remaining matrix. This step is repeated until none of the columns (or rows) have equal totals, or until further reduction will result in a matrix of size less than 3 x 3. This is the minimum size required for a two-dimensional solution. At this point, the reduced matrix is used as the input for MDS analysis, through which the appropriate number of dimensions is determined before constructing a conceptual map. As such, the chances of producing a two-dimensional map are much higher compared to the conventional procedure. Once the conceptual map is constructed, it is examined for patterns and labeled. Labels represent new requirements that the sorters could not easily group in one pile. These are difficult to extract by the conventional procedure.

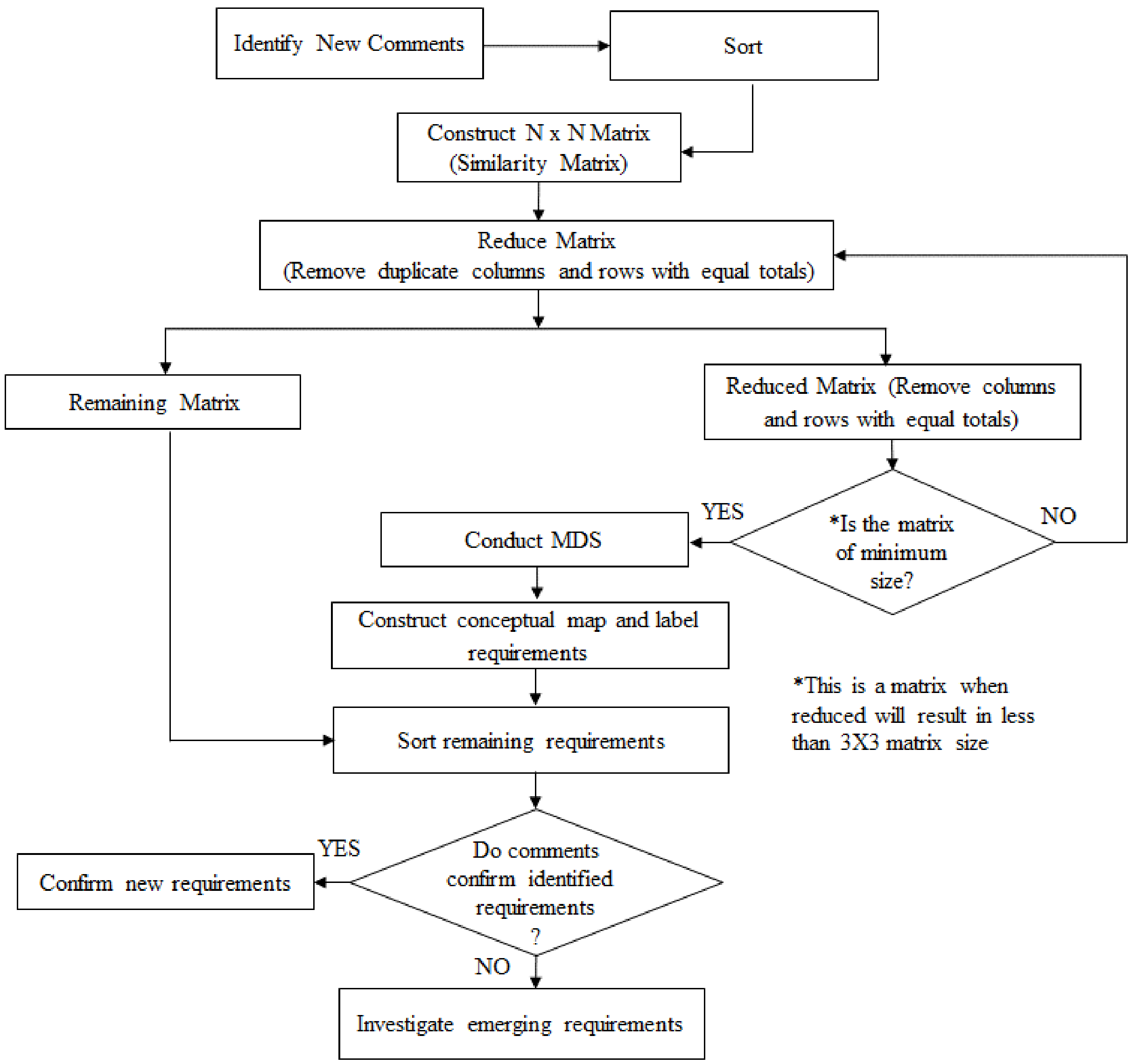


Figure 1. The Proposed Framework

The framework requires that all comments from the remaining matrix are resorted under the labels (requirements) identified. This task may be performed by the researcher or a group of sorters familiar with the product or service being evaluated. Comments that match the labels offer a confirmation of the requirements. Whereas, comments that do not match any of the labels represent emerging requirements to be investigated in future studies.

The following section represents an illustrative application of the proposed framework and the results obtained.

4. Illustrative Application

As an illustration, the data from Attar and

Weheba (2015) is reexamined following the framework shown in Figure 1. The original study was conducted to update the design of a students' satisfaction survey. The survey was designed to measure students' perceptions of in-class representations of selected publications. A total of ten questions were included, of which three were open-ended. Six questions were used to measure perceptions regarding key requirements including organization of the slides, manner of the presentation, providing illustrative examples, appropriate use of graphics, time management, and ability to answer questions.

Three open-ended questions were also included to solicit students' feedback on ways to improve their experiences. The survey was administrated 23 times during fall semester of 2013 and 2014 following in-class presentations. Students' responses to the open-ended questions were collected resulting in the 30 units (comments) considered in this research and shown in Appendix A.

The conventional CM procedure was applied and two new requirements were identified based on the clusters shown in Figure 2.

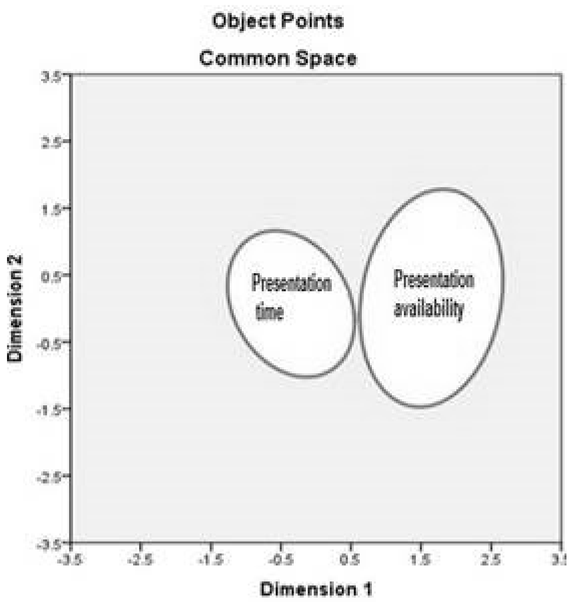


Figure 2: Clusters of student's requirements

Source: Attar and Weheba (2015)

Further examination of the scatter revealed some hidden requirements that were forced to fit a two-dimensional space. A three dimensional representation of the scatter is shown in Figure 3(a). Such a configuration suggested that a three-dimensional MDS representation may reveal some hidden requirements. Three projections were prepared using SPSS as shown in Figure 3(b). A projection of the scatter in the Y-Z plan suggests a different structure, and supports the need for another representation.

In this application of the new framework, five sorters from the class were asked to pile similar comments together. The five resulting matrices were combined in a 30X30 similarity matrix, part of which is shown in Figure 4.

This matrix was gradually reduced by removing duplicate rows and columns with equal totals. The 7 x 7 matrix shown in Figure 5 was found to be of minimum size. The remaining 23 comments were separated in a 23 x 23 matrix.

Next, SPSS was used to conduct MDS and the scree plot shown in Figure 6 supported a two-dimensional representation of the data. Hence a conceptual map was generated using two dimensions as shown in Figure 7. Three clusters were identified and labeled as presentation time, relevance and coverage.

The researcher was able to sort all of the remaining 23 comments under the three piles, except two comments. These are "Include applications for the research" and "Record presentations using Panopto." A group discussion with the sorters indicated some important facts. First, it appears that the students prefer papers with application examples as they provide better understanding of the concepts represented. In order to improve the level of student satisfaction, the presenters should spend more time researching the literature for applications and provide citations to interested students for more information.

Members of the discussion group agreed that including application examples is one aspect of the coverage requirement as was identified.

Second, the discussion revealed that the Panopto software license was recently acquired by the university and was available on the course management tool.

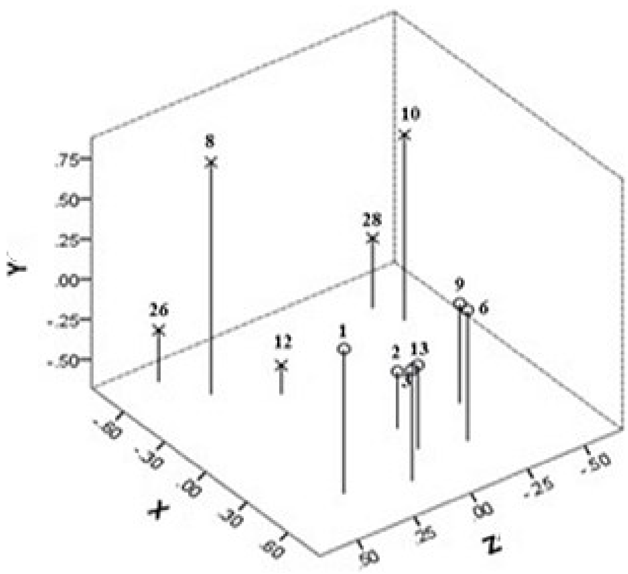


Figure 3 (a). A three-dimensional representation of the 30 comments

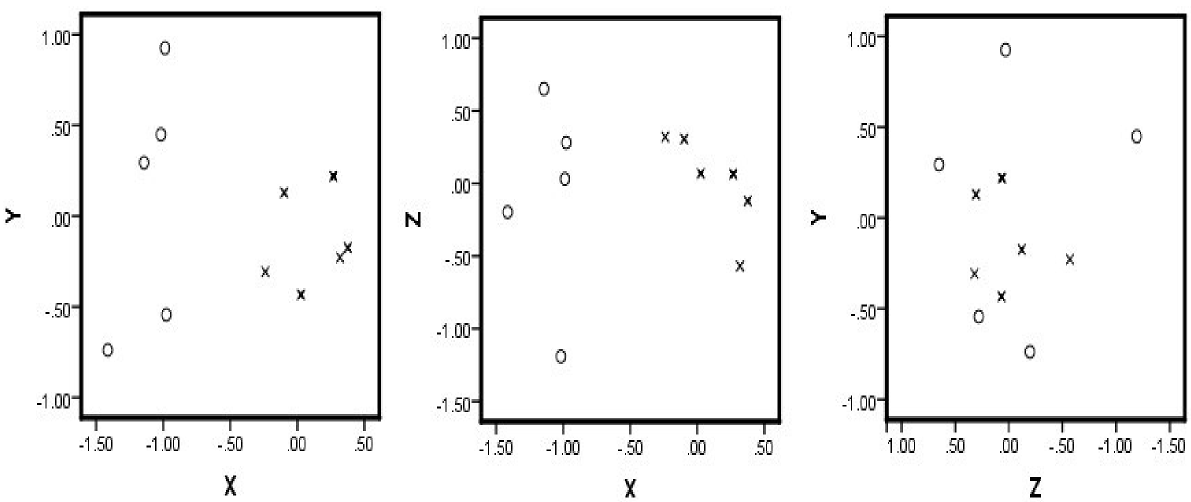


Figure 3 (b): Three different projections of the scatter

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1	5	3	4	4	5	3	5	0	4	1
S2	3	5	4	4	3	2	3	0	3	1
S3	4	4	5	5	4	2	4	0	3	2
S4	4	4	5	5	4	2	4	0	3	2
S5	5	3	4	4	5	3	5	0	4	1
S6	3	2	2	2	3	5	3	1	4	0
S7	5	3	4	4	5	3	5	0	4	1
S8	0	0	0	0	0	1	0	5	0	2
S9	4	3	3	3	4	4	4	0	5	0
S10	1	1	2	2	1	0	1	2	0	5
S11	4	3	3	3	4	4	4	0	5	0

Figure 4. Partial representation of 30 X 30 Similarity Matrix

	S1	S3	S8	S9	S10	S26	S28
S1	5	4	0	4	1	0	1
S3	4	5	0	3	2	0	2
S8	0	0	5	0	2	1	0
S9	4	3	0	5	0	0	0
S10	1	2	2	0	5	0	2
S26	0	0	1	0	0	5	2
S28	1	2	0	0	2	2	5
Totals	**15**	**16**	**8**	**12**	**12**	**8**	**12**

Figure 5. The reduced matrix (7x7)

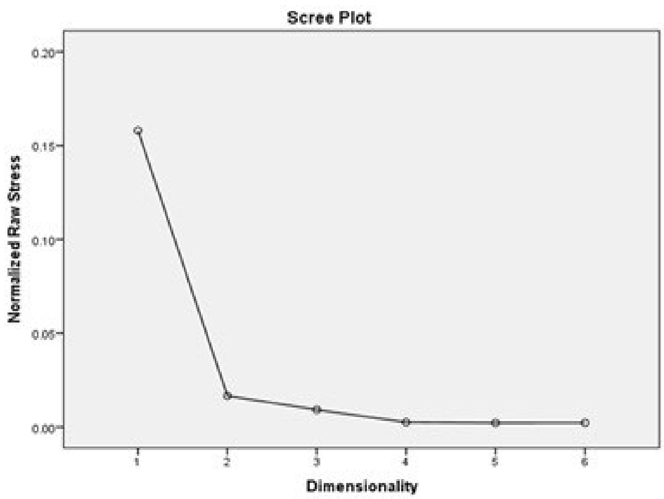


Figure 6. MDS Scree Plot Based on Reduced Matrix

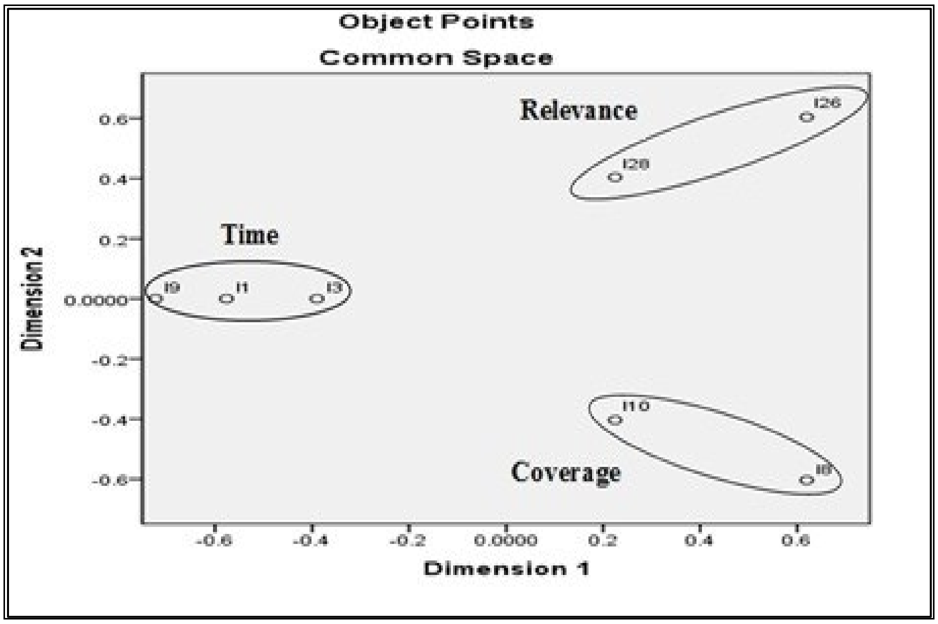


Figure 7. Conceptual Map from Reduced Matrix

Few instructors were able to utilize the software to record selected lectures. Students had positive perceptions of the tool as they were able to watch recorded lectures and review their notes at more convenient times. This new requirement is expected to affect student satisfaction as more instructors use the software and more students receive its benefits. This represented an emerging requirement to be confirmed in future studies.

These results as compared to what was reported by Attar and Weheba (2015) support the value of the new framework and highlight its advantages over the conventional approach.

7. Conclusions

The literature revealed that MDS has been extensively utilized to summarize qualitative data in order to help researches identify general structures. Given the difficulties encountered in interpreting structures in more than two dimensions, researches are often tempted to represent scatters with considerably fewer dimensions. Excessive reductions in dimensionality (at this stage) may hide new requirements or cause them to falsely merge with others. In this paper, we propose a new framework for performing MDS with the objective of discovering new requirements.

These are expected to be reflected in only few comments and likely to be hidden in conventional applications. The size of the similarity matrix depends on the number of statements (units of analysis) to be analyzed. In studies involving a large number of units, it may be more beneficial to reduce the size of the similarity matrix before preparing the MDS map. In other words, the researcher should seek a matrix representation with considerably smaller size. This would increase the chances that rare units (representing new requirements) are represented on the MDS map. To safeguard against excessive reduction, all remaining units are considered in a later stage to either confirm requirements or form new clusters. It is our hope that the new framework would help product and process designers as well as quality practitioners gain a better understanding of customers' requirements and their emerging needs.

References

Attar, M and Weheba, G (2015). Dynamic Validation of Customer Satisfaction Surveys, the Journal of Management and Engineering Integration, 8(1), 57-65

Borg, I and P. Groenen (2005) Modern Multidimensional Scaling: Theory and Applications, 2nd Edition, Springer Science + Business Media, Inc. NY, New York.

Carroll, J. D., & Wish, M. (1976). Multidimensional scaling: models, methods and relevance to DELPHI. In H. A. Linstone & M. Turoff (Eds.), The Delphi Method: Techniques and Applications , Addison Wesley.

Coxon, A. P. M., & Davies, P. M. (1982).*The user's guide to multidimensional scaling: with special reference to the MDS (X) library of computer programs*. London: Heinemann Educational Books.

Erickson, P. I., & Kaplan, C. P. (2000). Maximizing qualitative responses about smoking in structured interviews. Qualitative Health Research, 10, 829-840

Godfrey, A. B. (1999). Juran's quality handbook.

Hayes, B. E. (1997). Measuring customer satisfaction: survey design, use, and statistical analysis methods. ASQ Quality Press.

Jackson, K. M., & Trochim, W. M. (2002). Concept mapping as an alternative approach for the analysis of open-ended survey responses. Organizational Research Methods, 5(4), 307-336.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*(Vol. 11). Sage.

Malhotra, N. K., & Birks, D. F. (2007). *Marketing research: An applied approach*. Pearson Education.

Miles, M., & Huberman, M. (1994). Qualitative data analysis: An expanded sourcebook (2nd ed.). Thousand Oaks, CA: Sage.

Pothas, A.-M., Andries, D.,&DeWet, J. (2001). Customer satisfaction:Keeping tabs on the is- sues that matter. Total Quality Management, 12, 83-94.

Shinkareva, S. V., Wang, J., & Wedell, D. H. (2013). Examining similarity structure: multidimensional scaling and related approaches in neuroimaging. Computational and mathematical methods in medicine.

Sproull, N. (1988). Handbook of research methods: A guide for practitioners and students in the social sciences (2nd ed.). Lanham, MD: Scarecrow Press.

SPSS (2011), IBM SPSS Statistics for Windows, Version 20.0. IBM Corp. Armonk, NY.

Appendix A:

List of students' comments

S1	Presented the paper on time
S2	Focus more on the applications
S3	Covered material in paper in good time
S4	The paper was presented briefly
S5	Did not exceed the time
S6	Too long but understandable
S7	Time management
S8	Number of students in group
S9	It was a long presentation
S10	Everyone did so well
S11	It was lengthy
S12	Record presentation using Panopto
S13	Presentation was long
S14	Timing and easy decipher to the audience
S15	Was bit lengthy
S16	Time management and distribution
S17	Time usage, more explanations
S18	Presentation time and manner
S19	Finish the presentation as soon as possible
S20	Too lengthy and most of the topics were vague
S21	Time management
S22	The time was appropriate
S23	It was not long
S24	The presentation time management
S25	Too long
S26	Access to PowerPoints slides
S27	The length of the presentation
S28	We need verbatim form the text
S29	Too long presentation
S30	Time management and table explanation

Improving Supply Chain of Vehicle Distribution and Inventory System of a Car Manufacturer in Brazil

André Chia Ching Lin, Gustavo Pereira Doná, Hevelyne Duarte de Souza and R. Radharamanan
Mercer University, Macon, GA 31207 USA
radharaman_r@mercer.edu

Abstract

Supply chain management (SCM) is the oversight of materials, information, and finances as they move in a process from supplier to manufacturer to wholesaler to retailer to consumer. Supply chain management involves coordinating and integrating these flows both within and among companies. The ultimate goal of any effective supply chain management system is to reduce inventory with the assumption that products are available when needed. Supply chain management flows can be divided into three main flows: the product flow; the information flow; and the finance flow. The product flow includes the movement of goods from a supplier to a customer, as well as any customer returns or service needs. The information flow involves transmitting orders and updating the status of delivery. The financial flow consists of credit terms, payment schedules, and consignment and title ownership arrangements.

In this study, the data on the distribution of cars from a Brazilian production facility to twenty retail stores within that region were used to study the distributions and create simulation models in ARENA software. The statistical data gathered from the simulation were analyzed using Minitab and Excel. This analysis helped determine the current system capability to match the customer demand. Alternative distribution methods that are different from what is currently applied in the company were proposed, and compared whether these other methods would yield better results for the company. The comparative study resulted in an effective proposal for better distribution of cars and lower inventory levels at retail stores.

1. Introduction

Today, more than ever, organizations are trying to innovate and need to have more competitive advantages than their competitors to stay and grow in the market. To guarantee those advantages these organizations are seeking to develop and improve their supply chain practices which can be defined as a system that can be divided into parts. These parts consist of material suppliers, production facilities, chains of distribution, as well as the retail outlets. Supply chain management is what companies use to identify and minimize production costs and maximize their flexibility and competitive advantages.

However, due to the high complexity of supply chains, implementing and managing it can be very costly and time consuming. With that in mind, simulation and modeling can generate significant savings for the organization. Simulating a supply chain means to create a model with data that already exists with the objective to understand and improve the systems and processes within the organizations. It can also be used to identify new strategies and implement them in the organization. The modeling offers the management an opportunity to test and evaluate the proposed project before actually implementing it, thus, reducing the risk and cost for project implementation.

Several studies have been carried out on the production systems of the automotive industry, industrial relations, and the reasons for which a certain country or company is or is not competitively successful (Corrêa & Miranda, 1998). At present, automotive industries are operating in a growing and more competitive environment. In answer to competitive pressures and to consumer market demands, the companies are undergoing organizational changes, re-analyzing their activities and establishing new practices, such as focalization, globalization and a new standard of relationship with their commercial partners – suppliers and customers (Jones et al., 1997; Law, 1991). With a growth scenario found in few places in the world, the Brazilian automotive market has become attractive. It is currently the seventh largest producer and has the fourth largest consumer market, with 26 automakers established in the country, including the production of buses and trucks. From 2002 to 2011 the Brazilian automotive market grew 145%, while production progressed at a slower pace, reaching 109% in the same period. This growth scenario was found in very few places in the world, which made the Brazilian automotive market attractive for investment by automakers in the country, jumping from just over $ 1 billion in 2005 to nearly $ 5.4 billion in 2012 (ANFAVEA, 2013). Thus, the Brazilian automotive industry has a guaranteed presence in macroeconomic discussions of the country, with a share of the Brazilian industrial GDP of 21% and total GDP of 5%. In an environment where internal demand was higher than local production was capable to supply, the quest for competitiveness and productivity gains became a recurring theme in the executive agenda in order to increase market share and profitability, and at the same time the search for risk reduction and its consequent losses began to be discussed in the companies. The supply chain of the automotive industry is characterized by the following: a lot of links from raw materials to the final consumer (dealerships, automakers, auto parts manufacturers, manufacturing industries and raw material producers); outsourcing; increasing reliance on suppliers in the strategic procurement process; use of information technology and communication tools in a cooperative way; and globalization (Narasimhan & Talluri, 2009). Guedes et al (2015) provided a vision of the supply chain risk management in the Brazilian automotive industry through a case study, when samples of a strong (an automaker) and a weak (a second tier supplier) links were compared based on existing theory in order to understand their limits, variables, and potential new findings.

This study has an objective to improve the car distribution process performance from the manufacturer to retailers of one of the Brazilian automotive industries. For this study, the data on the distribution of cars from a Brazilian production facility of the Chevrolet division of General Motors Corp., to twenty retail stores that are within the region of the production facility were used. This provided necessary data to study the distributions and create a simulated program in ARENA software. The statistical data gathered from the simulation were analyzed using Minitab software. This analysis can help determine the current system capability to match the consumer demand. Through the course of the project, alternative distribution methods that are different from what is currently applied in the company were proposed, and compared whether these other methods would yield better results for the company. Hence, the expectation is that this comparative study will result in an effective proposal for better distribution for cars to retailers in the future and lower inventory levels.

2. Background research

Supply chain management is a network of facilities that produce raw materials, transform them into intermediate goods and then final products, and deliver the products to customers through a distribution system (Lee & Billington, 1995). The basic objective of supply chain management is to "optimize performance of the chain to add as much value as possible for the least cost possible". In other words, it aims to link all the supply chain agents to jointly cooperate within the firm as a way to maximize productivity in the supply chain and deliver the most benefits to all related parties (Finch, 2006).

Gunasekaran and McGaughey (2003) extended the scope of SCM beyond material management, partnership, information technology to the Total Quality Management areas like management commitment, organizational structure, training and behavioral issues. As firms' survival lies on integration, a good understanding of the integration process is a key aspect in SCM. Mouritsen et al (2003) discussed that basic hypothesis "the more integration (wider the scope) – the better the management of the chain" is not always true and proved that it depends very much on the "environment" of the supply chain and the power relations between the participants in the supply chain. Authors proposed a set of management techniques and tools to analyze successful SCM strategies. It is also observed that research is not limited to hypothesis testing and data analysis, but more advanced techniques like simulation, Artificial Neural Network (ANN), and Fuzzy logic are also used for optimization and decision making in SCM. Koh and Tan (2006) used the principles of fuzzy logic for analyzing and monitoring performance of suppliers based on the criteria of product quality and delivery time where as Chiu and Lin (2004) showed how the concepts of collaborative agents and artificial neural networks (ANNs) can work together to enable collaborative supply chain planning (SCP). Mentzer et al (2001) defined supply chain as systematic strategic coordination of the company in the framework of a company's traditional function as well as its supply chain, with the goal of improving in the long-term company's individual performance and of its the supply chain as a whole. Handfield and Nichols Jr. (2008) argued that supply chain embraces all areas related to merchandise flow and transformation from raw material extraction until the final users, as well as its respective information flows. Therefore, supply chain management endeavors a sustainable competitive advantage by integrating all these activities and improving its supply chain relationship.

Ballou (2004) stated that the supply chain involves multiple stakeholders, including suppliers, manufacturers, wholesalers, and retailers. Supply chain is a complex optimization problem that leads with system flows and capacity utilizations and its outcome is the result of a complex interaction between all of the factors. Ballou (2004) also indicated that, when managing a supply chain, inventory became too complex for mathematical methods, and it is interesting to apply the computer simulation. Terzi and Cavalieri (2004) complemented that the complexity, enormity, and the broader scope of a supply chain problem, weaken the analytical modeling approach, and the simulation modeling became a useful and satisfying method.

Russel and Taylor (1998) noted that it is interesting to apply simulation even for problems that have clear analytical solution methods, since simulation offers an easy way for problem testing and experimenting. Nerseian and Swartz (1996) systemized the simulation application in logistics by introducing methods to decide issues separately, the timing and quantity of orders, the level of inventory, and the number of warehouses.

According to Schunk and Plott (2000) simulation can be very effectively solved when applied to solving supply chain because of its capability of managing variability. Anderson and Morrice (1999) reaffirmed that supply chain can be an ideal tool when dealing with demand because it can handle a very complex dynamic system. However, according to Ingalls and Kasales (1999) simulating a supply chain can vary demand task since the model should simulate several key processes of these complex systems. Hieta (1998) complemented that, in fact, it is essential that the decision maker has an understanding of the random nature of demand and supply chain dynamic in order to interpret the results of the simulation.

In this paper, an attempt has been made to study a supply chain distribution of a Brazilian production facility of the Chevrolet division of General Motors Corp., using simulation. Through the course of the project, alternative distribution methods have been proposed in order to improve the company's car distribution process performance. Statistical study and analyses were applied to results interpretation in order to better access the random nature of a supply chain dynamic.

3. Methodology

3.1 Data from Chevrolet Brazil

Supply chain and management are different for each company, manufacturer, and retail store. The relationship between the manufacturer and the retail stores varies a lot and depends on many variables, like how big is the demand of the retail store, or how long is the lead time from the manufacturer to the retail store, and on what the final customer is waiting and willing to buy.

The automobile manufacturer of this study has factories throughout Brazil. Each plant is responsible to supply a certain number of retailers. This plant provides five different car models: Celta, Cruze, Cobalt, Onix, and Prisma to twenty different retailers.

The data for the model Celta is shown in Table 1. From this table, one can perceive that most of the retail stores data distributions follow the beta distribution.

The Beta distribution is a family of continuous probability distribution that is defined on an interval [0, 1] with two parameters, denoted by α and β. The beta function B is defined by:

$$B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt \tag{1}$$

Where: α > 0 and β > 0. The beta distribution can also be written as follows using the gamma function.

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{2}$$

Table 1. Celta model data

Retail Store	Demand	Stock (Inventory)	Production	Time in transit (days)
A	11.5 + 7 * BETA(0.441, 0.509)	5.5 + 6 * BETA(0.885, 0.991)	10.5 + 8 * BETA(0.418, 0.369)	3
B	17.5 + 8 * BETA(0.708, 0.832)	POIS(13)	15.5 + 10 * BETA(0.178, 0.107)	3
C	17.5 + 18 * BETA(0.198, 0.294)	TRIA(10.5, 13.3, 14.5)	19.5 + 13 * BETA(0.51, 0.709)	5
D	18.5 + 27 * BETA(0.186, 0.223)	12.5 + 16 * BETA(0.589, 0.375)	14.5 + 34 * BETA(0.304, 0.448)	6
E	17.5 + 93 * BETA(0.21, 0.138)	13.5 + 54 * BETA(0.431, 0.308)	11.5 + 86 * BETA(0.18, 0.148)	2
F	UNIF(4.5, 8.5)	-0.5 + 5 * BETA(1.1, 0.715)	3.5 + 5 * BETA(0.672, 0.672)	3
G	UNIF(68, 178)	54.5 + 61 * BETA(0.267, 0.226)	74.5 + 86 * BETA(0.0822, 0.074)	3
H	10.5 + 18 * BETA(0.244, 0.156)	2.5 + 11 * BETA(0.651, 0.359)	15.5 + 20 * BETA(0.141, 0.223)	3
I	44.5 + 75 * BETA(0.12, 0.152)	20.5 + 48 * BETA(0.137, 0.088)	61.5 + 35 * BETA(0.105, 0.153)	2
J	37.5 + 57 * BETA(0.114, 0.0757)	6.5 + 44 * BETA(0.467, 0.346)	43.5 + 43 * BETA(0.142, 0.18)	4
K	57.5 + 21 *	32.5 + 23 *	49.5 + 29 *	5

	BETA(0.176, 0.213)	BETA(0.31, 0.324)	BETA(0.165, 0.149)	
L	47.5 + LOGN(5.48, 11.3)	31.5 + 22 * BETA(0.171, 0.214)	39.5 + 22 * BETA(0.00559, 0.00445)	5
M	24.5 + 10 * BETA(0.466, 0.527)	15.5 + 12 * BETA(0.395, 0.335)	19.5 + 15 * BETA(0.538, 0.585)	8
N	8.5 + 10 * BETA(0.356, 0.394)	5.5 + 7 * BETA(0.859, 0.779)	10.5 + 8 * BETA(0.46, 0.46)	2
O	44.5 + 94 * BETA(0.191, 0.161)	29.5 + 42 * BETA(0.445, 0.318)	27.5 + 96 * BETA(0.182, 0.159)	3
P	96.5 + 92 * BETA(0.264, 0.359)	61.5 + 64 * BETA(0.141, 0.173)	109 + 60 * BETA(0.308, 0.408)	3
Q	5.5 + 11 * BETA(0.738, 0.566)	1.5 + 6 * BETA(0.866, 0.66)	8.5 + 7 * BETA(0.826, 0.794)	4
R	21.5 + 34 * BETA(0.232, 0.225)	16.5 + 20 * BETA(0.0908, 0.0954)	30.5 + 19 * BETA(0.296, 0.296)	2
S	34.5 + 21 * BETA(0.469, 0.417)	20.5 + 17 * BETA(0.0975, 0.077)	34.5 + 21 * BETA(0.165, 0.19)	4
T	14.5 + 12 * BETA(0.535, 0.712)	9.5 + 9 * BETA(0.65, 0.799)	13.5 + 12 * BETA(0.602, 0.722)	3

The expected value for the Beta distributed is:

$$E(x) = \mu = \frac{\alpha}{\alpha+\beta} \quad (3)$$

and the variance is

$$Variance(x) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \quad (4)$$

The distributions from the manufacturer to the twenty retail stores were gathered are later applied as inputs to the simulation (Lin et al, 2015).

3.2 Simulation

The distribution flowchart (Figure 1) highlights the manufacturer and its distribution to the retail stores, where "RS" stands for the Retail Store.

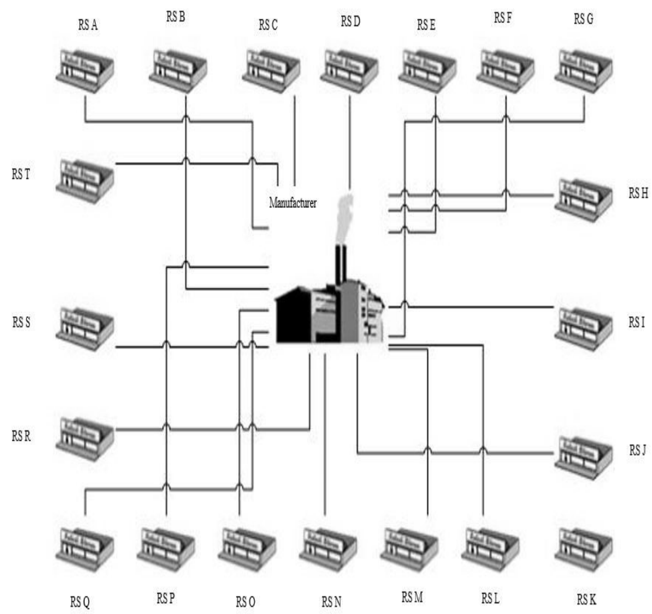


Figure 1. Distribution flowchart

The manufacturer produces a certain number of cars every month based on the past month's sales and distributes this number of cars of some different models for the retailers depending on how big the retailer is and its demand for each model of cars. Equation (5) defines the supplying priority DS_i (Days of Supply, it is the number of days that the retail store still have inventory to supply the demand) for a specific car, and for each retailer i.

$$DS_i = \frac{Pi+Si+Ti}{WSAi} \quad (5)$$

Every month, the supplier assembles P brand new cars and for each retailer i, it delivers P_i cars on the following month. The sum of all P_i values equals the number of total cars P. Besides that, each retailer, i, has an actual stock, S_i, and an average of monthly sales, WSA_i, and a T_i, that represents the number of cars that left the manufacturer and are in transit to the retail store.

The lowest DS_i determines the sequence of distribution, the retail store with the smallest DS will receive a new car, and so on. The manufacturer uses that formula to sort the retailers that would need to be supplied first, and thus to schedule a distribution plan to all P cars according to the urgency of each retailer. This process repeats monthly. When all the cars have been sequentially assigned to each retailer for the month, the process repeats itself for the next month. The old stock S_i and the monthly sales average, WSA_i, are updated. With this information, all DS's are calculated again and a new distribution schedule is planned.

To study and improve the current system of distribution from the manufacturer to the twenty retail stores, data, including the inventory of the retail stores, their demand, the number of cars that the manufacturer produces to each store every month, as well as how long it takes for the cars to go from the manufacturer to the stores were gathered and inserted them into ARENA.

However, there are two concerns in terms of the supplying priority DS number:

1) The period between DS update may affect the distribution outcome.

Currently the company operates with a monthly DS update. That is, the company assigns each retailer priority every month and sequences their delivery according to it. With that in mind, the system was simulated for different cases. First case, the company was simulated with a shorter period of DS update time, with a weekly DS update instead of a monthly update. In the second case, instead of associating the DS number with a time, the DS number will be updated for every item unit delivered or produced.

2) When and after which process should DS number be updated may also affect the outcome.

Another concern is in which part of the process DS priority number should be updated. Currently, the company updates their DS as soon as the product is being sent to the retailer. With that in mind, cases were simulated where the DS priority number will only be updated when the product actually arrived to the retailer.

Considering these two concerns, **six different scenarios** were created and the results of them were compared for each car model. Here, **(t)** stands for transit model, which means considering product in transit in the formula, and **(i)** stands for inventory model, which means the DS formula only updates when the retailer receives the product.

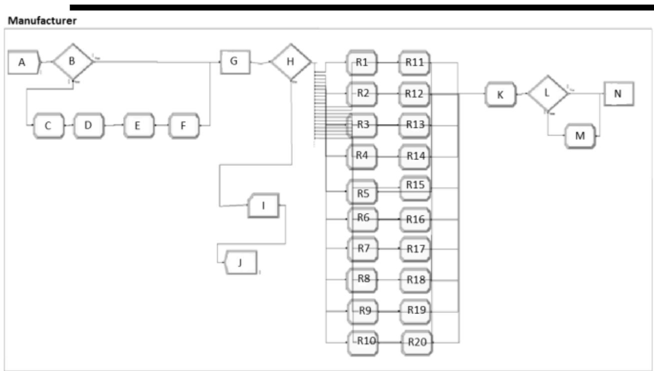
1) **(t.month)**: DS is updated monthly, and the product in transit is recognized in the formula (company's current operation).
2) **(i.month)**: DS is updated monthly, and recognized only when the retailer receives it.
3) **(t.week)**: DS is updated weekly, and the product in transit is recognized in the formula.
4) **(i.week)**: DS is updated weekly, and recognized only when the retailer receives it.
5) **(t.unit)**: DS is updated by each item/unit released, and the product in transit is recognized in the formula.
6) **(i.unit)**: DS is updated by each item/unit released, and recognized only when the retailer receives it.

In this phase, an attempt has been made to investigate how these assumptions may affect the simulation outcome (Lin et al, 2015).

4. Results and discussions

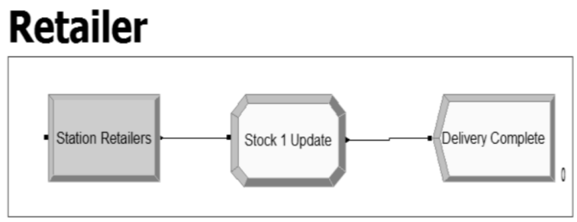
4.1 Simulation Results

The program on ARENA was divided into four parts. The following Figures are based on the original model of distribution. Figure 2 shows the cars leaving process from the manufacturer to the retail stores. Figure 3 shows the retail stores model. Figure 4 shows the retail stores and their customers' demand. The cycle reset model is shown in Figure 5.



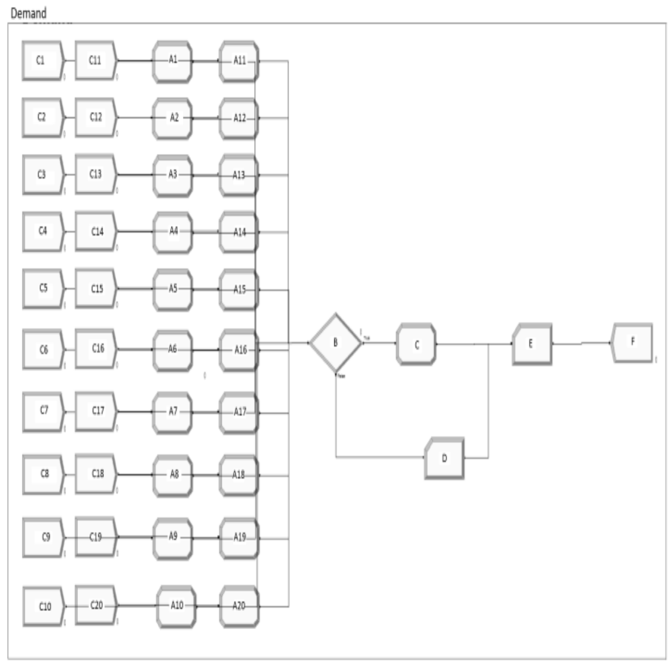
Where:
A: Create Cars: It represents the cars leaving the manufacturer
B: Decide if first month: In the first month this block decides to update all the distributions
C: Initial Production: Assign the distributions for the production (# of cars that the manufacturer is producing for each retail store)
D: Initial Inventory Level: Assign the number of cars in the beginning of the month based on the inventory for each retail store
E: WSA (Sales Average) Assign: Assign the distribution for the demand for each retail store
F: Assign DS Value: Assign the distribution for the DS priority for each retail store
G: Manufacturer: Station to the decision module
H: Decide Retailer Model: Decides which retail store has the lowest DS priority number
I: Excess: It counts the number of cars that the manufacturer produced even after all the retail stores have received their cars
J: Dispose Excess: End of simulation
K: Assign Counter: It increases the number of cars for each retail store
L: Decide if Complete: It decides if the retail stores received all the cars assigned to them
M: Assign Complete: It assign that the retail stores have received the cars
N: Route Retailer Station
R1 to R20: Assign the number to retail stores

Figure 2. Manufacturer model



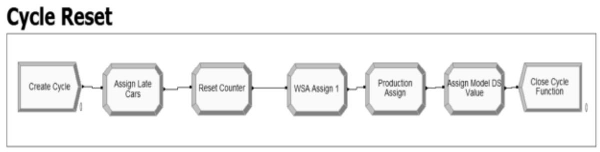
Where:
Station Retailers: Cars arrive to the retail stores
Stock 1 Update: It increases one car in the inventory for the stores
Delivery Complete: End of simulation

Figure 3. Retailer model



Where:
B: Decide Inventory: it decides if the retail stores have cars in their inventory to sell
C: Inventory Sales: it decreases
D: Record Customer without Cars: it counts the number of customers that leave the store without buying a car
E: Record Inventory: it counts the number of cars in the inventory on each store
F: Dispose Customer: End of simulation
C1 to C20: Creates the demand for each retail store
A1 to A20: Assign the customer to each retail store

Figure 4. Demand model



Where:
Create Cycle: Creates a new cycle every month or week
Assign Late Cars: Assign the number of cars that are late
Reset Counter: Transform the counter from the previous month or week to zero
WSA Assign 1: Assign a new demand for the stores
Production Assign: Assign a new number for the production
Assign Model DS Value: Assign a new DS number for the next month or week
Close Cycle Function: End of simulation

Figure 5. Cycle reset model

4.2 Analysis of Variance

The results of these different scenarios are collected and organized using the Minitab software. In order to test the validity of these two assumptions, ANOVA has been applied comparing if there is any difference between the transit and inventory method. It was assumed equal variances for the analysis and a significance level of 0.05; as well as a null hypothesis (meaning, if accepted, that all the means are equal) and an alternative hypothesis (meaning that at least one of the means is different). The results of the monthly, weekly, and per unit DS update were also compared.

One-Way ANOVA tests are being applied to identify if there is any statistical difference between each of these models. This method is used to test if there is any statistical difference between the mean of the samples. It assumes a null hypothesis if the compared samples have equal means; and an alternative hypothesis if the samples have different means. This way, one can identify if there is really a significant statistical difference between different scenarios and determine whether models yield better results.

The two assumption are being tested using this methodology: The study 1) compared the monthly, weekly, and unit DS update models; and 2) compared the inventory models (recognized only when the retailer received the product) and in transit models (product in transit are recognized in the formula).

The Analysis of Variance for Celta t.month, t.week, and t.unit are showed below.

```
Factor Information
Factor Levels Values
Factor 3 t.month, t.week, t.unit

Analysis of Variance
Source  DF   Adj SS   Adj MS  F-Value  P-Value
Factor   2  2282917  1141459    11.48    0.000
Error   57  5669710    99469
Total   59  7952627
```

From the ANOVA test, it is seen that the p-value is lower than 0.05. That is, the null hypothesis, all means are equal is not valid. The monthly, weekly, and per unit DS update have different means, that is **1) the period between DS update may affect the distribution outcome** may be true.

The Tukey test is also applied as a follow-up test to ANOVA. In Figure 6, when the interval does not contain zero, it means that there is a significant difference between two samples.

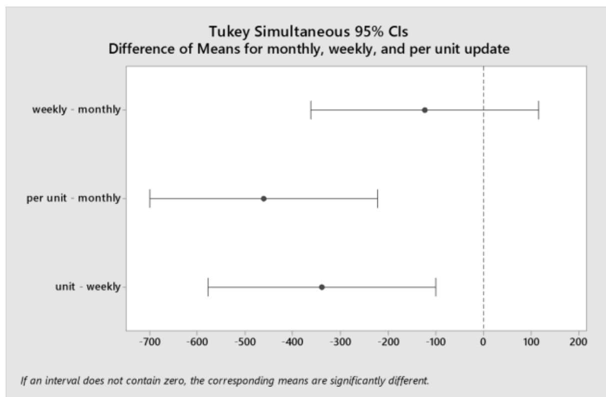


Figure 6.Tukey test for Celta t.month, t.week, and t.unit scenarios

From the graph, it can be concluded that the t.month, and t.week means are equal. Meanwhile, t.unit scenario is different from t.month and t.week. Similar results have been obtained from other four car models.

The Analysis of Variance for Celta t.month and i.month is also performed.

From this ANOVA test, it can be seen that the p-value is higher than 0.05. That is the null hypothesis, all means are equal is valid. The t.month and i.month have equal means, which is the assumption **2) when and after which process DS number should be updated may also affect the outcome** may be incorrect.

```
Factor Information
Factor  Levels  Values
Factor       2  t.month, i.month

Analysis of Variance
Source  DF   Adj SS  Adj MS  F-Value  P-Value
Factor   1       38      38     0.00    0.988
Error   38  6147624  161780
Total   39  6147662
```

The Tukey test is also applied as a follow-up test to ANOVA. From Figure 7, it is reaffirmed that the i.month and t.month have equal means. Similar results have been obtained from the other four car models.

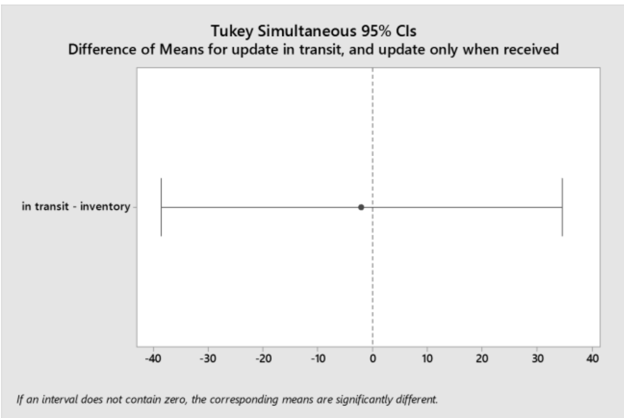


Figure 7. Tukey test for Celta t.month and i.month scenarios

4.3 Statistical analysis

The statistical data from the simulation are collected from the Rockwell Arena Software and then data of average inventory level and average customer backorder of each scenario are collected and analyzed using the Minitab software. Figure 8 shows the average inventory level for each scenario of Celta.

From the Boxplot (Figures 9), it can be concluded that there is no significant difference between the transit models **(t)** and inventory models **(i)**. But there are differences between the monthly, weekly and per unit models. The unit models demonstrated better performance compared to the other models. Similar results are obtained from other four car models.

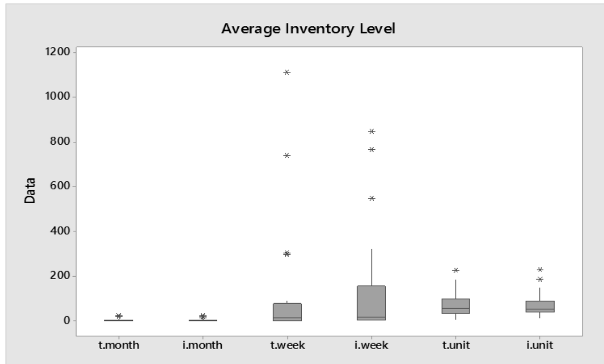


Figure 8. Celta average inventory level for each scenario

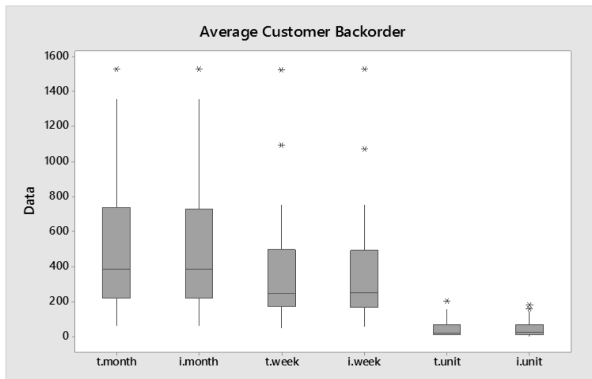


Figure 9. Celta average customer backorder for each scenario

5. Conclusions and Recommendations

5.1 Conclusions

Banks and Malave (1984) identified inventory control problems as one of the most frequent areas of application for simulation methodology. They propose six categories of simulation techniques usage in modeling and analyzing inventory systems:

(1) analytic solution impossible or analytic solution extremely complex, (2) comparison of models, (3) verification of analytic solutions, (4) variance reduction, (5) model validation and verification, and (6) optimization. A lot of simulation studies have been performed in order to handle uncertainty inherent in system operation environment. For instance, simulation is used to enhance operational system decision making in an uncertain environment (Petrovic et al., 1998). Chandra et al. (2001) investigated information coordination influence on the demand forecast accuracy in a supply chain through simulation. Landeghem and Vanmaele (2001) studied the behavior of the supply chain under different sources of uncertainty using the Monte Carlo simulation approach.

In this study, based on the simulation and statistical analyses performed, and by comparing the outcomes of different scenarios, it was possible to conclude that the assumption **1) "the period between DS update may affect the distribution outcome**" is true.

By reducing the period, the time between each DS update, it reduces the number of customer backorders. That is, the plant can better attend to the demand. As shown in the Figure 9, the scenarios 5-6 **(t.unit)** and **(i.unit)** demonstrated the best performance in attending to the demand.

It is known that the manufacturer used to update the DS number weekly, which is a better method than the current (monthly updates). It is recommended that the manufacturer updates its DS priority number every time a car leaves the manufacturer and arrives at the retail stores. Doing this, the manufacturer will have more control and will be fairer with the distribution between the retail stores.

However, there is no evidence to support the assumption **2) "when and after which process should DS number be updated may affect the outcome**". The data collected showed that there is no significant statistical difference between recognizing and not recognizing the product in transit in the DS priority formula.

The problem that this project was supposed to solve was to create a model of distribution that could reduce the inventory on the retail stores and increase the number of customers that leave the store with their new cars based on the current model that the manufacturer uses.

The model that had the best performance, which means, the model that presented a small inventory for the retail stores (in general) and more customers leaving the stores with their new cars was the model that considers calculating the DS number per unit, considering or not the cars that are in transit from the manufacturer to the stores. This model per unit calculates the DS priority number every time a car is produced, so it updates the DS number more frequently.

Implementing the proposed methodology will certainly help the management in satisfying the daily supply needs of the retail stores thereby meeting the demand and the customer satisfaction.

5.2 Recommendations

For further studies, it would be more accurate if the other manufacturers as well as all the models of cars that each manufacturer produces are considered in the simulation. Other limitations in this research were some assumptions made in the simulation model to make possible for the model to run and the reduced number of months of data that were given.

6. Acknowledgment

This project was funded by the Mercer Summer Engineering Experience (MeSEE) program of Mercer University School of Engineering during the Summer Semester of 2015.

References

Anderson, E. G., & Morrice, D. J. (1999). *A simulation model to study the dynamics in a service-oriented supply chain, In Proceedings of the 1999 Winter Simulation Conference, ed., P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, 742-748.* Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

ANFAVEA (2013), *Brazilian National Association of Automotive Vehicles Manufacturers, Statistical Yearbook of the Brazilian Automotive Industry (Anuário da Indústria Automobilística Brasileira)* , São Paulo.

Ballou, R. H. (2004). *Business Logistics/supply Chain Management: Planning, Organizing, and Controlling the Supply Chain,* Pearson PrenticeHall.

Banks J. and Malave C. O. (1984). *The Simulation of Inventory Systems: An Overview,* Simulation Councils, Inc., June, 1984, pp 283-290.

Chandra C., Chilov N. and Smirnov A. (2001). *Analysis of Supply Chain Bullwhip Effect Using Simulation Modeling,* CONSA Research Report Series "Simulation: Applications, Research and Education in the Baltic Area", Presentations at the Special CONSA Conference at Linkoping University, Nr 5, June, Linkoping University Sweden.

Chiu, M., & Lin, G. (2004). Collaborative Supply Chain Planning using the Artificial Neural Network Approach. *Journal of Manufacturing Technology Management,* 15 (8), 787–796.

Corrêa H. L., & Miranda, N. G. M. (1998). Supply network management in the Brazilian automotive industry, *Integrated Manufacturing Systems,* 9/5, 261–271.

Finch, B. J. (2006). *Operations Now: Profitability, Processes, Performance,* 2nd edn, McGraw-Hill/ Irwin, United States

Guedes, E. J. G., Bittar, A. V., Di Serio, L. C., & Oliveira, L. H. (2015). Risk Management in the Supply Chain of the Brazilian Automotive Industry, *Journal of Operations and Supply Chain Management,* Volume 8 Number 1, 72–8773.

Gunasekaran, A., & McGaughey, R. E. (2003). TQM is Supply Chain Management, *The TQM Magazine,* 15 (6), pp. 361–363

Handfield, R. B., & Nichols Jr., E. L. (2008). *Introduction to Supply Chain Management.* 2nd Edition, Prentice Hall.

Hieta, S. (1998). Supply chain simulation with Logsimsimulator, *In Proceedings of the 1998 Winter Simulation Conference Ed, D.T Medeiros, E.F. Watson, J.S Carson and M.S Manivannan, 323-326,.* Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Ingalls, R., & Kasales, C. (1999). CSCAT: The Compaq supply chain analysis tool, *In Proceedings of the 1999 Winter Simulation Conference, ed., P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, 1201-1206.* Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Jones, D. T., Hines, P. and Rich, N. (1997), "Lean logistics", *International Journal of Physical Distribution & Logistics Management,* Vol. 27 No. 3/4.

Koh, S. C. L., & Tan, K. H. (2006). Operational Intelligence Discovery and Knowledge-Mapping Approach in a Supply Network with Uncertainty, *Journal of Manufacturing Technology Management,* 17 (6), 687–699.

Law, C. M. (1991), *Restructuring the Global Automobile Industry: National and Regional Impacts,* Routledge, London.

Lee H. L., & Billington C. (1995). The Evolution of Supply Chain Management Models and Practice at *Hewlett-Packard, Interfaces,* 25(5) pp. 42-63.

Lin, A. C. C., Doná, G. P., & Souza, H. D. (2015). *Improving Supply Chain of Vehicle Distribution and Inventory*

System of a Car Manufacturer in Brazil, Report, MeSEE, pp. 1-92.

Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., D.Smith, C., & Zacharia, Z. G. (2001). Defining Supply Chain Management, *Jornal of Business Logistics,* (Vol. 22).

Mouritsen, J., Skjøtt-Larsen, T., Kotzab, H. (2003). Exploring the Contours of Supply Chain Management. *Integrated Manufacturing Systems,* 14 (8), pp.686–695.

Narasimhan, R., & Talluri, S. (2009). Perspectives on Risk Management in Supply Chains, *Journal of Operations Management,* 27(2): 114–118.

Nerseian, R., & Swartz, G. (1996). *Computer Simulation in Logistics,.* Quorum Books, (ISBN 0899309852).

Petrovic D., Roy R., & Petrovic R. (1998). Modeling and Simulation of a Supply Chain in an Uncertain Environment, *European Journal of Operational Research,* Vol. 109, pp 299-309.

Russel, R., & Taylor, B. (1998). *Operation Management: Focusing on Quality and Competitiveness.* New Jersey: Prentice Hall.

Schunk, B., & Plott, D. (2000). Using Simulation to Analyze Supply Chains, *In Proceedings of the 2000 Winter Simulation Conference .ed., J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 1095-1100.* Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Terzi, S., & Cavalieri, S. (2004). Simulation in the Supply Chain Context: A Survey, *Computers in Industry, 53, 3-16.*

Van Landeghem H. and Vanmaele H. (2001). *Robust Planning: A New Paradigm for Supply Chain Planning,* In: Proceedings of the 15th European Simulation Multiconference, June 6-9, pp 866-873.

Reliability Analysis on the Life of the Light Bulbs: A Case Study

Isadora V. Sousa and R. Radharamanan
Mercer University, Macon, GA 31207
radharaman_r@mercer.edu

Abstract

Reliability refers to the probability of a system or component to perform its required functions under stated conditions for a specified period of time. In life data analysis and accelerated life testing data analysis, as well as other testing activities, one of the primary objectives is to obtain a life distribution that describes the time-to-failure of a component, subassembly, assembly or system. This analysis is based on the time of successful operation or time-to-failure data of the item (component), either under use conditions or from accelerated life tests.

This paper analyzes the life of light bulbs in hours. Nonparametric techniques for ungrouped data were used to determine the appropriate distribution governing the data collected. The life data of the street light bulbs in hours collected in Brazil were used in this analysis. The first step in the analysis was to define the distributions to be tested and their respective probability density function (PDF) and cumulative distribution function (CDF). Sequentially, histograms and probability plots were drawn. From the plots, the coefficients of determination were determined, and based on their values the best fit distribution was selected. The final analysis of the reliability function plot and hazard function plot were made to verify the reliability of the chosen data distribution. The results were then compared to the published data available in the literature. The analysis of the results indicated that Weibull distribution best fits the data analyzed.

1. Introduction

In recent years, the demand for systems that perform better and cost less has suddenly become essential. The concomitant requirement to minimize the probability of failures has also become of primary concern in the development of new products and systems. In this context, the term reliability is defined as the probability that a system will perform its intended function for a specified period of time under a given set of conditions. The developed knowledge to analyze systems failures and minimize their occurrence can be applied to real life situations in which reliability considerations appear (Lloyd and Lipow, 1962; Zack, 1992; Lewis, 1994).

Reliability can be expressed as a function of time or as it is commonly known the inverse of the probability of failure.

$$R(t) = 1 - F(t) \quad (1)$$

Where $R(t)$ is the probability that a system operates without failure for a length of time t and $F(t)$ is the probability that failure takes place at a time less than or equal to *t*.

The four most commonly assumed failure probability distributions are (Holmberg and Folkeson, 1991; Hoyland and Rausand, 2009; and Modarres, 2009):

- Exponential distribution is time-to-failure distribution of chance failures; It is generally used for constant failure rate models.
- Normal distribution commonly referred to as Gaussian distribution has location and scale parameters and is the most used distribution in statistics. It is used for time-dependent failure rate models.
- Lognormal distribution is usually used to describe failures caused by fatigue, uncertainty in failure rates, and fatigue cracks. It is often used to model parts or components that fail primarily due to stress or fatigue.
- Weibull distribution has shape and scale parameters capable of describing premature, constant and wear out failures, which are the three basic elements of a typical bathtub curve. It is widely used for brittle materials in modeling their distribution of time-to-failure and of their strength.

In this paper, the life of lights bulbs in hours are analyzed using nonparametric and parametric methods. Histograms and probability plots were made assuming that the data follow either Normal, Lognormal, or Weibull distributions. The best fit distribution was selected based on the value of the coefficient of determination (R^2). The reliability function plot and the hazard function plot were made to verify the reliability of the chosen data distribution. The results were then compared to the published data available in the literature.

2. Background

Mirzai et al (2006) stated that failures of transformers in sub-transmission systems not only reduce reliability of power systems but also have significant effects on power quality since one of the important components of any system quality is reliability of that system. The failure modes of transformers can be represented by the Weibull distribution. Weibull statistics have been widely used and accepted as a successful mathematical method to predict the remaining life time of any equipment. Based on their study several useful conclusions were presented both for power systems operators and manufacturers for improving the reliability of transformers.

Protective relays are components of a protection system in a power system domain that provides decision making elements for correct protection and fault clearing operations. Failure of the protection devices may reduce the integrity and reliability of the power system protection that will impact the overall performance of the power system. Hence it is imperative for power utilities to assess the reliability of protective relays to assure it will perform its intended function without failure. Ridwan et al (2010) discussed the application of reliability analysis using statistical methods to assess and evaluate the reliability of numerical overcurrent protective relays from two different manufacturers.

In reliability analysis of car maintenance forecast and performance, researchers have mostly dealt with problems either without maintenance or with deterministic maintenance when no failure can occur. This can be unrealistic in practical settings. Owhor et al (2015) developed a statistical model to evaluate the effect of predictive and preventive maintenance schemes on car performance in the presence of system failures where the forecasting objective is to minimize schedule duration.

The extended utilization of light bulbs in most industrial, commercial, and domestic daily activities represents a potential application of reliability analysis. The choice of the lighting techniques - the existing ways to arrange lights - and the lamps - the light bulbs - for the rooms any environment will make a difference in comfort level, energy use, and cost.

Considering lighting as a single appliance for instance, it can be as much as 25 percent of a home's electricity consumption. Understanding the life of light bulbs and their operational reliability will decrease the probability of failure, extra costs, time wasted, and the psychological effect of inconvenience (Kuhn, 1971).

The reliability effects in almost all aspects related to the possession of a property: cost management, customer satisfaction, risk analysis, resources management and product safety, and quality are presented and discussed by several authors. Zacks (1992) divided reliability analysis according to the system's technological design and function. The mission reliability describes the successful operation of a device which is constructed for the performance of one mission only, and the operational reliability describes the successful operation of a system that is turned on and off intermittently to perform a certain function. There are systems whose operational reliability increase with lifetime, and some whose reliability decrease with lifetime. The operational reliability is a function of both the readiness and the probability of continuous functioning of the system for a specified period, and it is the best methodology to evaluate the performance of the life of light bulb data (Zacks, 1992).

Radharamanan (2010) carried out reliability analysis using time to failure data of incandescent light bulbs recorded at accelerated life conditions using statistical distributions. Radharamanan and Ansuj (2011) analyzed public transportation bus motor failures using both parametric and nonparametric techniques to determine the appropriate distribution governing these failures. They used six different parametric distributions and the coefficients of determination to find the best fit distribution for the data analyzed.

In the broadest sense, reliability is associated with dependability, successful processes, and absence of interruptions. A product or system fails when it ceases to perform its intended function. According to Lewis (1994), it is necessary to define failure quantitatively in order to take into account the more delicate forms of failure, through deterioration or instability of function. Berk (2009) in his book "System Failure Analysis" performed a fault-tree analysis to look for the causes at the point where the failure manifests itself or the immediately adjacent causes that can induce the command event considering that the system experiences a failure when the switch is actuated and the light bulb fails to illuminate. He pointed out five causes for the failure: open light bulb filament, fracture in the bulb glass, contaminated terminals in the socket, bulb not fully screwed into the socket, and no electrical energy from the socket. Based on Berk's (2009) conclusions regarding the reasons behind the light bulb failing to illuminate, the purpose of this paper is to use the coefficient of determination, the reliability function plot and the hazard function plot to find the best distribution for the life data of light bulbs in hours.

3. Methodology

The life data of light bulbs in hours were collected from 100 street bulbs in Brazil. Using the life data of the light bulbs, plots were generated to give a visual representation and to give insight into which probability distribution the data best fit. A description of the analytical tools and techniques used are as follows:

Statistical Software: Microsoft Excel® (2010) is a common data analysis computer application that can be very useful in the development of statistical analyses for large data sets. Excel was used throughout this study for the organization of raw data, calculation of descriptive data and the generation of all charts and plots.

Histogram: A histogram is a graphical demonstration of data using bars of different heights (looks similar to a bar chart). In particular, a frequency histogram is a special histogram that uses vertical columns to show how often value (or a given range) occurs in a sample.

Reliability Plotting: A reliability plot is a visual plot that demonstrates the change in reliability of a sample over time. A perfect system at time 0 would have an expected reliability of 1. As time increases, the expected reliability in the system should tend towards zero. For the ungrouped data:

$$\hat{R}(t_i) = \frac{N+1-i}{N+1} \tag{2}$$

where: *N* = is the sample size; *i* = 0, 1, 2, ..., N. Then the cumulative distribution function (CDF) is given by:

$$\hat{F}(t_i) = \frac{i}{N+1} = 1 - \hat{R}(t_i) \tag{3}$$

Cumulative Hazard Plotting: Similar to reliability plots; cumulative hazard plots visually examine distribution model assumptions for reliability data. Cumulative hazard plots consist of a plot of the cumulative hazard versus the time of the i^{th} failure. For the ungrouped data:

$$\hat{H}(t_i) = \ln(N+1) - \ln(N+1-i) \tag{4}$$

Probability Plots: Probability plots are visual ways of summarizing reliability data by plotting cumulative distribution function (CDF) estimates versus time while using a log-log scale. When doing this, the x-axis is time and the y-axis is a cumulative percentile. The three probability plots used in this study are: Weibull, Normal and Lognormal. All of the probability plots are formed and given a line of best fit that follows:

$$y = ax + b \tag{5}$$

The parameters *a* and *b* can be found using Microsoft Excel®. Once the parameters are obtained, additional, unique analysis is done for each distribution. The methodology for these distributions are given below.

Weibull Distribution plotting: Using a two parameter Weibull distribution, the plot will be defined by:

$$X\ axis = \ln t$$
$$Y\ axis = \ln \ln \left[\frac{1}{1-F(t)}\right]$$
$$\hat{\theta} = e^{\left(-b/a\right)} \tag{6}$$

where *a* and *b* are found by plotting a line of best fit.

Normal Distribution plotting: The normal distribution, unlike the Weibull, cannot be inverted to obtain y in analytical form. Thus, the linear equation is found by taking:

$$X\ axis = Failure\ measurement\ (in\ hours)$$
$$Y\ Axis = \phi^{-1}(F)$$
$$\hat{\sigma} = 1/a \text{ and } \hat{\mu} = -b/a \tag{7}$$

Lognormal Distribution plotting: Lognormal probability plotting is similar to normal except the x-axis is plotted by the natural log of the failure criteria:

$$X\ axis = \ln t$$
$$Y\ Axis = \phi^{-1}(F)$$
$$\hat{\omega} = 1/a \text{ and } \hat{t}_0 = e^{\left(-b/a\right)} \tag{8}$$

Initially, a scatter plot (Figure 1) was drawn in order to visualize the data graphically. This plot facilitates the understanding and analysis of the raw data. To create the scatter plot the data were organized in ascending order and plotted with the light bulbs' lifetime in hours on the y-axis.

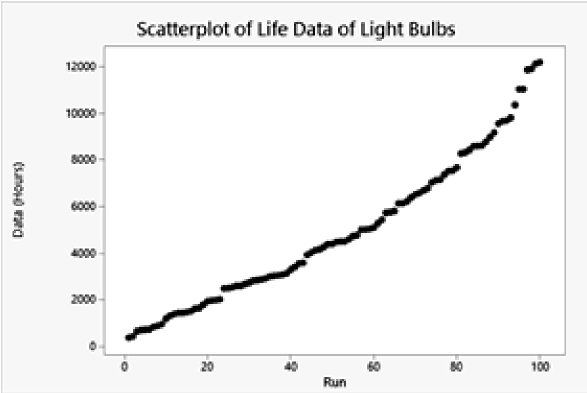


Figure 1. Scatter plot

To plot the histogram, the first step was to find the range of the data analyzed (i.e., the maximum minus the minimum value). Knowing the range, an interval width was chosen such that the data could be divided into some number N of groups. If too few intervals are used, the nature of the distribution is obscured by lack of resolution, and if a large number of intervals are used it would hide the nature of the distribution because of the great fluctuations in frequency. More data points allow a larger number of intervals to be used effectively, and result in better representation of the distribution. The number of intervals for the data was carefully chosen to obtain a better plot.

In order to specify the type of distribution that would be appropriate for the life data of the light bulbs, three different distributions were tested; they were Normal, Lognormal, and Weibull. To determine how well they fit the data, their corresponding probability plots were drawn and their coefficients of determination(R^2) were established. When the R^2 value is closer to one, the distribution fits the data better. The probability plotting is an extremely useful technique. With relatively small sample sizes it yields estimates of the distribution parameters and provide both a graphical picture and a quantitative estimate of how well the distribution fits the given data. Table 1 shows the probability graphing information used for each distribution (Lewis, 1994; and Modarres, 2009).

Table 1: Probability plotting information

Distribution	F(t)	y(F)	x(t)
Normal	$\Phi\left(\frac{t-\mu}{\sigma}\right)$	$\Phi^{-1}(F)$ *	T
Lognormal	$\Phi\left(\frac{1}{\omega}\ln\frac{t}{t_0}\right)$	$\Phi^{-1}(F)$ *	$\ln(t)$
Weibull	$1-e^{-\left(\frac{t}{\Theta}\right)^m}$	$\ln\left(\ln\left(\frac{1}{1-F}\right)\right)$	$\ln(t)$

* y = Φ^{-1}(F) = Excel formula NORMSINV

Continuous random variables find extensive application in reliability analysis and a significant number of standardized probability distributions is employed to model the behavior of these variables. The normal distribution is the most used in statistics, and it is frequently referred to as the Gaussian distribution. The normal distribution arises in many contexts, but can be a bad choice when the data exhibit a significant skewness. In this case, the lognormal distribution is indicated. The lognormal distribution is occasionally referred to as the Galton distribution and represents a continuous probability distribution of a random variable whose logarithm is normally distributed. The Weibull distribution is also widely used in reliability analysis for describing the distribution of times to failure and of strengths of brittle materials, such as ceramics.

To test the reliability of the system, nonparametric methods for ungrouped data were used. These are methods in which the reliability functions and hazard functions are plotted without an attempt to fit them to a particular distribution. The ungrouped data methods were chosen because they best fit the nature of the life data of the light bulbs. It provides the time at which the bulb fails to illuminate (Lewis, 1994).

4. Results and Discussions

The histogram for the life data of the light bulbs is shown in Figure 2. The analysis of the histogram plot indicates that the data analyzed do not follow a specific tendency and pattern. It also shows that the PDF for a normal distribution do not fit the life data of the light bulbs.

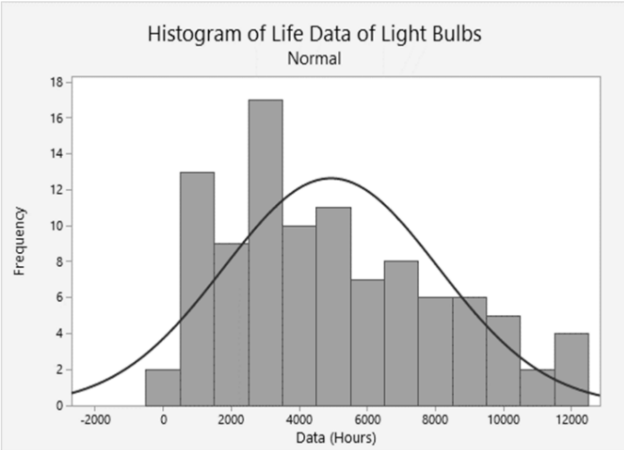


Figure 2. Histogram plot

The reliability and hazard function plots for the life data of the light bulbs in hours are shown in Figures 3 and 4 respectively.

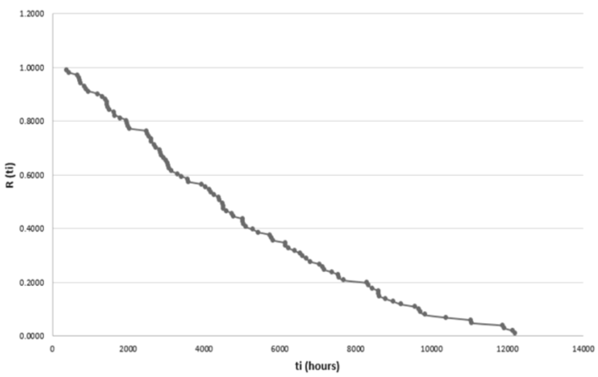


Figure 3. Reliability function plot

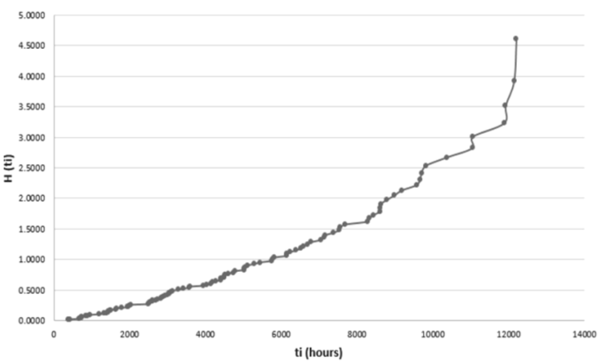


Figure 4. Hazard function plot

The reliability function plot shows that the reliability of the light bulbs decreases with increasing time of use. Moreover, the concave upward behavior of the hazard function provides evidence of an increasing failure rate as the life time of the light bulbs increases.

Figure 5, 6, and 7 show the cumulative distribution function plots for the Normal, Lognormal and Weibull distributions respectively. To analyze the close approximation of the distribution to the given data, the trendline was drawn following the cumulative distribution function of the initial life data of the light bulbs. Tables 2and 3 show the distribution parameters (R^2 and P-value) and the fitted equations for the Normal, Lognormal, and Weibull distributions.

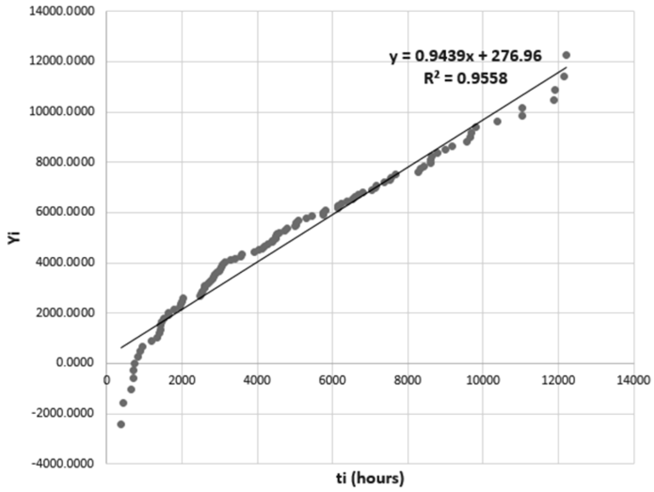


Figure 5. Normal probability plot

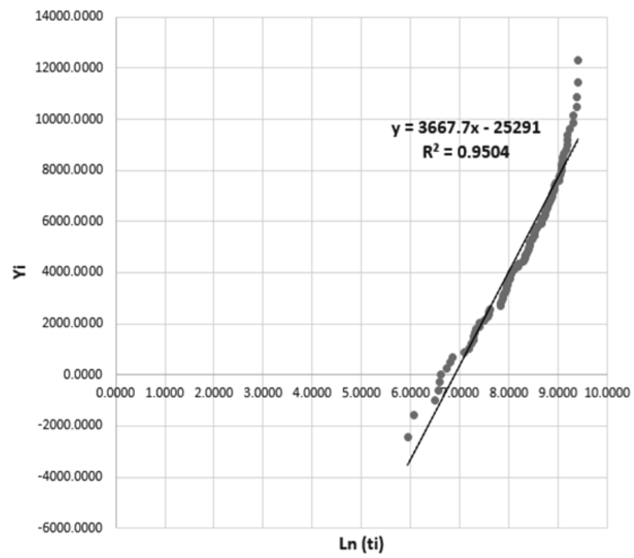


Figure 6. Lognormal probability plot

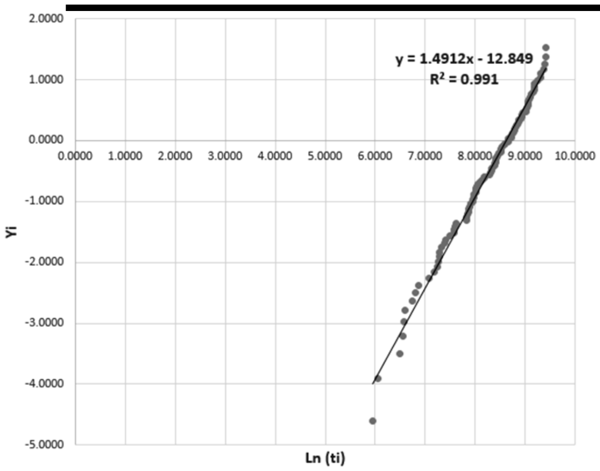


Figure 7. Weibull probability plot

Table 2. Distribution parameters

Distribution	R^2	P-value
Normal	0.9558	< 0.01
Lognormal	0.9504	< 0.01
Weibull	0.9910	> 0.15

Table 3. Fitted equations

Distribution	Fitted Equation
Normal	Y = 0.9439 X + 276.96
Lognormal	Y = 3667.7 X - 25291
Weibull	Y = 1.4912 X - 12.849

Analyzing the plots above, the Weibull distribution was found to be the best fit for the life data of the light bulbs with R^2 value close to one (0.991) and the highest *P*-value (> 0.15). It means that the observed data are consistent with the assumption that the null hypothesis is true, and thus that hypothesis must be accepted. The Normal and Lognormal plots show low *P* values, and high R^2 values. This means that the observed data are inconsistent with the assumption that the null hypothesis is true, and thus that hypothesis must not be accepted.

5. Conclusions

The analysis of both parametric and nonparametric methods found that the Weibull distribution was the best fit for the life data of the light bulbs. It also shows that the light bulbs reliability decreases when the time of use increases; consequently, the failure rate increases when the life time of the light bulbs increases.

The life of a light bulb in hours can vary depending on the conditions in which the light works. A system experiences a failure when the switch is actuated and the light bulbs fail to illuminate as shown in the book "System Failure Analysis" in which Berk (2009) performed a fault-tree analysis to look for the causes of failure when it occurs by itself or the immediately adjacent causes that can induce the interruption. Based on Berk's findings, there are five causes for the light bulb failure: open light bulb filament, fracture in the bulb glass, contaminated terminals in the socket, bulb not fully screwed into the socket, and no electrical energy from the socket.

Although this article does not analyze further the causes of the failure and ways to avoid it, there are different devices used to increase the life in hours of a light bulb or reduce its failure rate. Kuhn (1971) in his article "Rectifier for extending the life of light bulbs" presented a lighting accessory device for rectifying the alternating current voltage input to a conventional light bulb. Berlin, Mensing and Nilssen (1990) in their article "Lamp life extender" presented a thin disc-like device for attachment to the base end of an incandescent light bulb to provide an auxiliary voltage reducing circuit interfaced between the central base contacts of the light bulb and its cooperating lamp holder for purposes of extending the rated life of the light bulb.

References

Berk, J. (2009). *Systems Failure Analysis*. AMS International, Russell Township, Ohio.

Berlin, D. M. Jr., Mensing, C. E., Nilssen, O. K. (1990). *Lamp Life Extender.* Abstract, United States Patent, US 4229680 A.

Holmberg, K., and Folkeson, A. (1991).*Operational Reliability and Systematic Maintenance.* Elsevier

Science Publishers Co., Inc., New York.

Hoyland, A., and Rausand,M. (2009). *System Reliability Theory: Models and Statistical Methods.* John Wiley & Sons, Inc. New York, NY, p. 275 – 294.

Kuhn, R. D. (1971). *Rectifier for Extending the Life of Light Bulbs*. Abstract, United States Patent, US3617766 A.

Lewis, E. E. (1994). *Introduction to Reliability Engineering.* John Wiley & Sons, Inc. Canada.

Lloyd, D. K., and Lipow, M. (1962). *Reliability: Management, Methods, and Mathematics.* Prentice-Hall, Inc., Englewood Cliffs, NJ.

Microsoft Excel (2010). Microsoft Office Professional Plus. [Computer Software]. Santa Rosa, CA: Microsoft Corporation.

Mirzai M., Gholami, A., and Aminifar, F. (2006). Failures Analysis and Reliability Calculation for Power Transformers. *J. of Electrical Systems-(on-line),* 2-1: pp. 1-12.

Modarres, M., Kaminskiy, M., and Krivtsov, V. (2009). *Reliability Engineering and Risk Analysis: A Practical Guide.* 2nd Edn., Taylor & Francis, Inc.

Radharamanan, R. (2010). Reliability Analysis on Incandescent Light Bulbs: A Case Study. *Proc. Int. Conf. on Manufacturing and Engineering Systems (ICMES 2010),* Dec. 16-18, Southern Taiwan University, Taiwan, pp. 23-30.

Radharamanan, R. and Ansuj A. P. (2011). "Reliability Analysis on a Public Transportation Bus Motor Failure" *The Journal of Management and Engineering Integration,* Vol. 3, No. 2, Page 74-81.

Ridwan, M. I., Yen, K. L., Musa, A., and Yunus, B. (2010). Application of Life Data Analysis for the Reliability Assessment of Numerical Overcurrent Relays. *Int. J. of Electrical, Computer, Energetic, Electronic and Communication Engineering,* Vol: 4, No: 12, pp. 1854-1860.

Owhor, S. C., Gambo, A. A. I., Ojo, V. K., and Daniel, D. (2015). Reliability Analysis of Car Maintenance Forecast and Performance. *American Journal of Engineering Research (AJER),* V4, Issue 7, 290-299.

Zacks, S. (1992). *Introduction to Reliability Analysis.* Springer-Verlag, Inc., New York, NY p.135.

A Fault-Tolerant Tree-Based Multicast Routing Algorithm for 2D Mesh Multicomputers

Masoud E. Shaheen
Fayoum University, Egypt
masoud.shaheen@fayoum.edu.eg

Abstract

Distributed memory systems are the most favorable architectures used in advanced research problems. 2D mesh networks are popular architectures that have been implemented in many distributed memory systems. In this paper, a new fault tolerant tree-based multicast algorithm for wormhole routed 2D mesh multicomputers is introduced. It works perfectly for the most commonly encountered faults in mesh networks, f-rings and f-chains. The proposed algorithm is proved to be deadlock free. Performance results of a simulation study on different 2D mesh networks are discussed to compare the proposed algorithm with a previous algorithm.

1. Introduction

2D Mesh networks are popular architectures that have been implemented in Illiac IV, MPP, DAP, CM-2, and Intel Paragon with variation (Hwang, 1993). They are preferred because they offer useful edge connectivity, which can be used for I/O controllers and can be partitioned into submeshes, simplifying the design of routing algorithms that are independent of network size (Duato, 1993). Also, they have many desirable properties including scalability, low cross-section bandwidth, and fixed degree of nodes. Multicast is an essential communication technique in several network topologies especially in wireless mesh network (Matam and Tripathy, 2016).

Fault tolerance refers to the ability of the system to operate correctly in the presence of faults (Chang and Chiu, 2002). The current generation of routers is very robust and do not appear to fail very often in practice. However, in some environments it is very important that failures can be expected and addressed. These operating environments are characterized by differing component failure rates, and abilities for repair. This has led to the development of a range of approaches to fault tolerant routing in mesh networks.

Efficient routing algorithms are critical to the performance of multicomputers systems. An effective multicast routing must be deadlock free and should minimize communication latency and channel traffic. Many deadlock-routing algorithms have been developed for wormhole communications networks. An optimal message routing should achieve both minimum traffic and minimum latency for the communication patterns involved. Unfortunately, finding optimal message routing has been shown to be NP-hard for most common multicomputers topologies.

In this paper, a new **f**ault **t**olerant **t**ree-**b**ased **m**ulticast routing algorithm (FTTBM) for 2D meshes is introduced. FTTBM algorithm is a tree-based multicast algorithm, which attempts to deliver the message to all destinations in a single multi-head worm that splits at some routers and replicates the data on multiple output ports. FTTBM algorithm works perfectly for convex faults, f-rings and f-chains, which are the most commonly encountered faults in mesh networks. It requires only local knowledge of all faults and is effective even in the presence of a large number of faults. Also, the routing information that must be kept by each non-faulty node is small. FTTBM algorithm is compared with FT-cube2 algorithm (Chang and Chiu, 2002).

This paper is organized as follows. In section 2, a previous fault tolerant multicast routing algorithm in 2D meshes is briefly discussed. In section 3, FTTBM algorithm is presented. A performance analysis is introduced in section 4. Conclusions are introduced in section 5.

2. Previous work

In recent years, fault tolerant multicast routing in direct networks is considered an important issue in computer researches. In this section, another algorithm is discussed. A fault tolerant unicast routing algorithm, FT-cube2, in 2D meshes was proposed (Chang and Chiu, 2002). In FT-cube2 algorithm, the well-known e-cube routing algorithm enhanced in order to handle multiple fault regions in 2D meshes using two virtual channels. The e-cube algorithm routes a message first along the row and then along the column in a 2D mesh. The e-cube algorithm provides deadlock-free shortest path routing without requiring virtual channels. Also, when faults exist in a mesh, an e-cube hop of a message maybe blocked the misrouted message to avoid the fault region.

An abundant amount of research has been done on fault tolerant multicast routing in mesh networks (Boppana and Chalasani, 1995; Park et al, 2000; Shaheen and Abukmail, 2012; Xiang et al, 2009; Xiang, 2011; Khan and Goodridge, 2015; Pfeiffenberger et al, 2015). Among them, this paper will briefly explain the most relevant research to this algorithm. An effective technique for fault tolerant multicast routing to develop the current wormhole switched routing algorithm was presented (Boppana and Chalasani, 1995). The algorithm measured randomly located faulty blocks and fault information that uses only local knowledge of faults and tolerates non-overlapping f-rings. Their fault tolerant scheme uses these f-rings to transmit messages around fault regions. Both f-rings and f-chains are allowed in their fault model, but must be on rectangle shape.

Two fault tolerant wormhole routing algorithm F3 and F4, whose tolerate concave fault regions using three and four virtual channels respectively was proposed (Park et al, 2000). These algorithms guarantee the delivery of messages between any pair of non-faulty and connected nodes in the network. A novel fault tolerant multicast routing algorithm, for wormhole routed 2D mesh multicomputer was presented (Shaheen and Abukmail, 2012). This algorithm is a unicast/tree based multicast routing algorithm. The proposed routing algorithm, called fault tolerant deadlock-free multicast, FTDM, works perfectly for the most common faults in 2D mesh networks, f-rings and f-chains. This algorithm is proved to be deadlock free.

A new deadlock-free adaptive routing scheme for 3D meshes using only two virtual channels per physical channel by building full use of the idle channels was proposed (Xiang et al, 2009). This new deadlock-free adaptive routing scheme is also extended to n-D meshes and derived from the planar network fault model. Also, a deadlock-free routing scheme for meshes derived from a new virtual network partitioning scheme, called channel overlapping was introduced (Xiang, 2011). According to their new virtual network partitioning scheme there are two virtual networks that can share some common virtual channels. This fault tolerant deadlock-free adaptive routing algorithm is also extended to the one in an n-D mesh using two virtual channels per physical channel. In this algorithm fault blocks are constructed inside divide planes, which turn on a lot of globally unsafe fault free nodes.

3. The proposed algorithm

All fault tolerant routing algorithms which was proposed recently concentrate on unicast-based multicast algorithms. In unicast-based multicast algorithms, a source node sends messages to its set of destinations by sending a sequence of separate unicast messages to each destination. Unicast-based algorithms require a large number of startups to send a message to a large set of destinations. Also, they are inefficient because they allow a message to be delivered to only one destination, which leads to multicast operations being implemented as multiple phases of multicast message exchange. So, contention freedom must be guaranteed not only among the worms of a given phase, but also among worms in different phases.

In this section, a new fault tolerant routing algorithm, FTTBM, for 2D meshes is introduced. FTTBM algorithm is a tree-based multicast algorithm, which attempts to deliver the message to all destinations in a single multi-head worm that splits at some routers and replicates the data on multiple output ports. Tree-based multicasting techniques offer very promising means of achieving extremely efficient multicast routing.

FTTBM algorithm considers convex faults which are the most commonly encountered faults in mesh networks. A convex fault is a fault region such that there is a rectangle whose interior nodes contains the faulty components of the fault region and all processors and links on its four boundaries are fault-free. A fault ring (f-ring) consists of the fault-free nodes and channels that are adjacent to one or more components of the associated fault region. If a fault region includes boundary nodes, the fault ring reduces to a fault chain (f-chain). FTTBM algorithm requires only local knowledge of all faults.

To define the path routing functions, which determines the next node for which the path of FTTBM algorithm will be visited, some definitions are introduced:

1) Let $f_{bi} = (x_{bi}, y_{bi})$, and $f_{ei} = (x_{ei}, y_{ei})$.
2) The fault region number *i*, F_i, is described by two nodes, f_{bi}, f_{ei}, where f_{bi} is located in the west south corner of the fault region, while f_{ei} is located in the east north corner of the fault region i.e.,
   $F_i = \{(x, y): x_{bi} < x < x_{ei} \wedge y_{bi} < y < y_{ei}\}$.
3) Width of a fault region F_i is defined as follow: $d_{Fi} = | x_{ei} - x_{bi} |$
4) The variable d_x is equal to 1 if the direction of the message path is from west to east or -1 if it is from east to west.
5) **L**ast **b**oundary **n**ode (*LBN*) is the label of last boundary node of a fault region which the message path visits it. The value of LBN is zero if the message path is in a non-fault region, while it is non-zero if the message path is in a fault region.

3.1. Routing Functions

FTTBM algorithm assigns a label for each node based on the position of that node in a Hamiltonian path. The Hamiltonian path in a network is an undirected path that visits every node in a graph exactly once where the first node in path is labeled 1 and the last node in the path is labeled N, where N is the network size (Mohapatra, 1998). The label assignment function Q for an m × n 2D mesh using a Hamiltonian path can be expressed in terms of the x- and y-coordinates of nodes as follows:

$$Q(p_i) = Q(x_i, y_i) = \begin{cases} y_i \times n + x_i + 1 & y_i \text{ is even number} \\ y_i \times n + n - x_i & y_i \text{ is odd number} \end{cases}$$

FTTBM algorithm creates the routing decision at each sending node. The path followed by a message is defined by one of the two routing functions. Each function is defined as a function of the node currently holding a message, and the destination node of this message. The function returns a neighboring node of the current node to which the message must be forwarded. Let c be a current node, and d is a destination node.

1) The first routing function used in FTTBM algorithm is defined as:

R (c, d) = w, where

$$Q(w)=\begin{cases} \max\{Q(z) : Q(z) \leq Q(u) \text{ and } z \text{ is a neighboring node of } c\} & \text{if } Q(c) < Q(d) \\ \min\{Q(z) : Q(z) \geq Q(u) \text{ and } z \text{ is a neighboring node of } c\} & \text{if } Q(c) > Q(d) \end{cases}$$

It was proved that for two arbitrary nodes c and u in a 2D mesh, the path selected by the routing function R is the shortest path between them (Lin and Ni, 1991). This routing function is deadlock-free even using the path based on facility. FTTBM algorithm which uses the routing function R in each region does not contain any fault nodes.

2) The second routing function used in FTTBM algorithm at a fault region Fi is defined as:

R′(c, d) = w, where

$$w=\begin{cases} (x_c, y_c-1) & \text{if } x_d = x_c \quad \text{(i.e., c and d on the same column of } F_i) \\ (x_c, y_c+1) \text{ if } x_d = x_c + d_x \times d_{Fi} & \text{(i.e., c and d on different columns of } F_i) \\ (x_c + d_x, y_c) & \text{otherwise} \quad \text{(i.e., c and d on the same row of } F_i) \end{cases}$$

FTTBM algorithm uses the routing function R' in fault regions only. Figure 1 illustrates the different cases of the routing function R' and the way of its work around the fault region. The direction of the message path may be from west to east, Figure 1(A) or from east to west, Figure 1(B). It is clear that, the path selected by the routing function R' is the shortest path between the two nodes c and u. Also, it is clear that, the routing function R' is deadlock-free, because it works on three boundaries only of each fault region, i.e., the cycle is not complete.

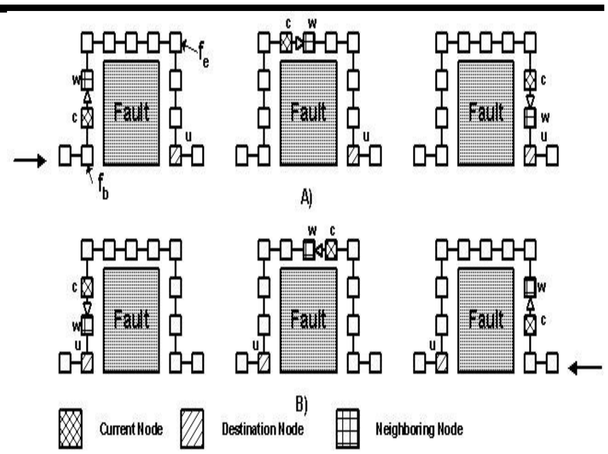


Figure 1 The routing path using R'

3.2. The algorithm structure

FTTBM algorithm divides the network into two subnetworks. The high-channel subnetwork contains all of the channels whose direction is from lower-labeled nodes to higher-labeled nodes, and the low-channel subnetwork contains all of the channels whose direction is from higher-labeled nodes to lower-labeled nodes. At the source node, FTTBM algorithm divides the 2D mesh into two submeshes, MU and ML, where every node in MU has a higher label than that of the source node and every node in ML has a lower label than that of the source node. Also, FTTBM algorithm divides set of destinations, D, into two subsets, DU and DL, where DU contains the destination nodes in MU and DL contains the destination nodes in ML. The messages will be sent from the source node to the nodes in DU using the high-channel network and to the destination nodes in DL using the low-channel network. The routing technique used by FTTBM decreases the packet transmission time involved between the source and the destination nodes.

It is assumed that each node in the network knows all fault nodes. Upon receiving the message, each current node determines whether it is the first destination node. If so, it is removed from the destination nodes and receives the message. At this point, if set of the destination nodes are not empty, the FTTBM algorithm continues according to the following method:

The current node determines the next node for which the path of FTTBM algorithm will be visited according to the routing function R. If the next node is a non-fault node, then the current node sends the message to it. Otherwise, the current node determines the next node according to the routing function R'. The current node checks LBN value, if it is equal to zero, then the current node divides the destination set into two subsets and sends the message twice to the next node. The header of the first copy contains the destinations behind the fault region, while the header of the second copy contains the remaining destination nodes. After this current node, the message path is branched into two paths, the first path may circle around the fault region while the other may continue the original path. If LBN value is not zero, then the current node sends the message to the next node without division.

Algorithm FTTBM
Input: The label of last boundary node *LBN*, destination set D, and the message *mg*.
Output: Compute the next node of the message path and send a copy (or two copies) of the message to it.
Procedure:

[1] if $c = d_1$ then
1. 1) remove c from D
1. 2) RCV(c, *mg*)
[2] if $D = \phi$ then stop
[3] $w = R(c,d_1)$
[4] **if** $w \in$ *Non-Fault-Nodes* then
SND (w, 0, D, *mg*)
else
4. 1) $D_2 = D$
4. 2) if $LBN \neq 0$ then $u = Q^{-1}(LBN)$
if y_c is *even* then $d_x = 1$ Else $d_x = -1$
$u = (x_c+d_x*d_{Fi}, y_c)$
if $d_x = 1$ then $D_2(c) = \{v: v \in D \wedge x_v \geq x_c \wedge y_v \leq y_{ei}\}$
else $D_2(c) = \{v: v \in D \wedge x_v \leq x_{ei} \wedge y_v \leq y_{ei}\}$
$D_1 = D - D_2(c)$
$w = R'(c, u)$
$LBN = Q(u)$
if $D_1 \neq \phi$ then SND (w, 0, D_1, *mg*)
end if
4. 3) $w = R'(c, u)$
4. 4) if $D_2 \neq \phi$ then SND (w, *LBN*, D_2, *mg*)

End FTTBM

FTTBM algorithm uses two functions. The first is SND (w, LBN, D, mg), where w is the next node of the path, LBN is the label of last boundary node, D is set the destinations, and mg is the message body. It is used to pass the message from the current node to the next node of the path. The second function is RCV (c, mg) makes the node c receives the message.

4. A performance analysis

To study the performance of FTTBM algorithm, it will be compared with one of fault-free unicast routing algorithms, FT-cube2 (Chang and Chiu, 2002). To perform this comparison, two essential performance metrics in direct networks are calculated. In subsection 4.1, the communication latency and channel traffic are calculated for FTTBM and FT-cube2 algorithms. In subsection 4.2, the results and discussions are introduced.

4.1. Communication latency and channel traffic analysis

The communication latency is the greatest number of channels which the message takes to reach its destinations. The channel traffic is the total number of channels used to deliver the message to all destinations. They affect the overall performance of the multicomputer system and the granularity of parallelism that can be exploited from the system (Mohapatra, 1998).

Now, the communication latency and channel traffic are calculated for FTTBM and FT-cube2 algorithms. The following formulas can be used to calculate the communication latency for FTTBM algorithm, MLat.

1) F is number of fault regions
2) $dist(d_i, d_{i-1}) = |x_{di} - x_{di-1}| + |y_{di} - y_{di-1}|$
3) $Lat(D) = \sum_{i=1}^{|D|} dist(d_i, d_{i-1})$
4) $D_i = \{(x, y) : (x, y) \in D \wedge x > x_{bni} \wedge y < y_{ei}\}$
5) $D_{bni} = \{(x, y) : (x, y) \in D \wedge Q(x, y) < Q(bn_i)\}$

6) $DL_{bni} = D_{bni} \cup D_i - \bigcup_{j=0}^{i-1} D_i$
7) $Latf_i = Lat(DL_{bni}) + 2\,|y_{ei} - y_{bni}|$
8) $Latf = \max\{Latf_i,\ 1 \le i \le F\}$

The communication latency of FTTBM algorithm, MLat, is given by:

$$MLat = \max(Lat(D), Latf) \qquad (1)$$

The channel traffic of FTTBM algorithm, MTraf, is given by:

$$MTraf = Lat(D) + \sum_{i=1}^{|F|} Latf_i \qquad (2)$$

The communication latency of FT-cube2 algorithm, FLat, is given by:

$$FLat = \max\{ Flat_i,\ 1 \le i \le |D| \} \qquad (3)$$

Where $Flat_i = |x_{di} - s_x| + |y_{di} - s_y| + 2 * |y_{ei} - y_{bi}|$

The channel traffic of FT-cube2 algorithm, FTraf, is given by:

$$FTraf = \sum_{i=1}^{|D|} Flat_i \qquad (4)$$

Lemma: FTTBM algorithm is deadlock-free
Proof: There are two cases as following:
Case 1: Nonfault regions

Because a cyclic dependency among resources is a necessary condition for deadlock, since a message is routed at any node according to routing function R, which proved deadlock-free (Lin et al, 1994), and monotonic order of requested channels is guaranteed, therefore, a cycle cannot exist within this path in the network; hence no cyclic dependency can be created among the channels.

Case 2: Fault regions

Since a message is routed at any faulty nodes according to routing function R´, and a message never visits an f-ring and f-chain more than twice (at most as a row message and once as a column message), then, a cycle cannot exist within this path in the network. Hence, FTTBM algorithm is proved deadlock-free.

4.2. Results and discussions

To compare the performance of the FTTBM algorithm and FT-cube2 routing, up to 30 random 2D mesh networks that contained, double channels was used. The two algorithms were written using c++ language and was implemented on a PC. The networks were generated with different values for R × C, where R is the number of rows ranging between 10 and 27, and C is the number of columns ranging between 10 and 40. The notation F is used to represent the number of fault regions. This configuration creates different networks with a number of processors ranging from 100 to 1080. The average number of destinations in each row is ranging from 1 to 10.

The equations from 1 to 4 are used to calculate communication latency and channel traffic for both algorithms in the different networks. Figures from figure 2 to figure 7 show the results. The continuous line represents results of FTTBM algorithm, while the dotted line represents results of FT-cube2 algorithm.

Figure 2 plots the communication latency for various values of average number of destinations, ranging from 1 to 20. The figures show that, the communication latency computed by FTTBM algorithm increases as the number of destinations increases. The rate of the increasing, decreases as a number of destinations increases because the destinations are converged. The communication latency computed by FT-cube2 algorithm is nearly constant as the number of destinations increases. This is because FTTBM algorithm is a multicast routing while FT-cube2 algorithm is a unicast routing.

Figure 3 plots the communication latency for various sizes of one fault region, ranging from 1 to 100 where RxC=1x1 to 10x10 and |D| is equal to 10. The figure shows that, the communication latency computed by FTTBM algorithm decreases as size of the fault region increases, while the communication latency computed by FT-cube2 algorithm increases. This is because the path computed by FT-cube2 algorithm circles around the fault region. So, when size of the fault region increases, number of used channels increases. Clearly, at small fault region sizes, the communication latency computed by FT-cube2 algorithm is less than that computed by FTTBM algorithm while at large fault region sizes, the communication latency computed by FTTBM algorithm is less than that computed by FT-cube2 algorithm.

Figure 4 plots the communication latency for various values of number of fault regions, ranging from 1 to 10 and |D| is equal to 15. The figure shows that, the communication latency computed by FTTBM algorithm decreases as number of fault regions increases, while the communication latency computed by FT-cube2 algorithm increases. Clearly, at a small number of fault regions, the communication latency computed by FT-cube2 algorithm is less than that computed by FTTBM algorithm while at large number of fault regions, the communication latency computed by FTTBM algorithm is less than that computed by FT-cube2 algorithm.

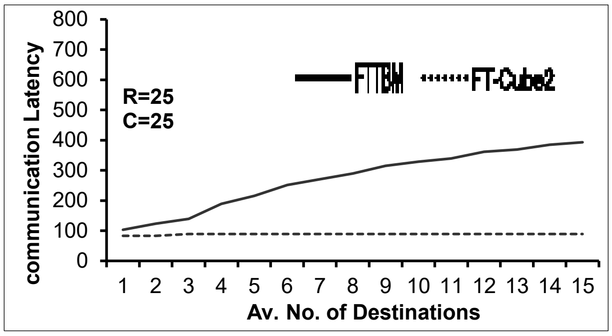


Figure 2 Communication Latency Vs. Av. No. of Destinations

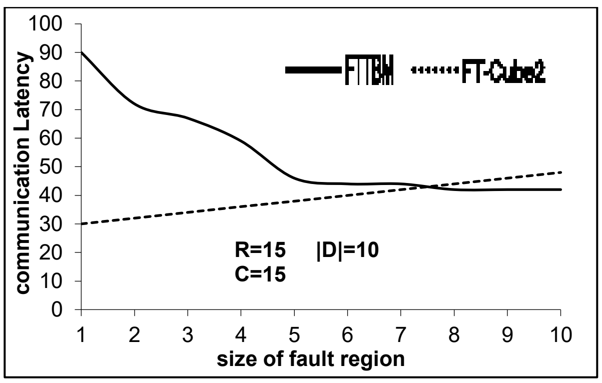


Figure 3 Communication Latency Vs. Size of the fault region

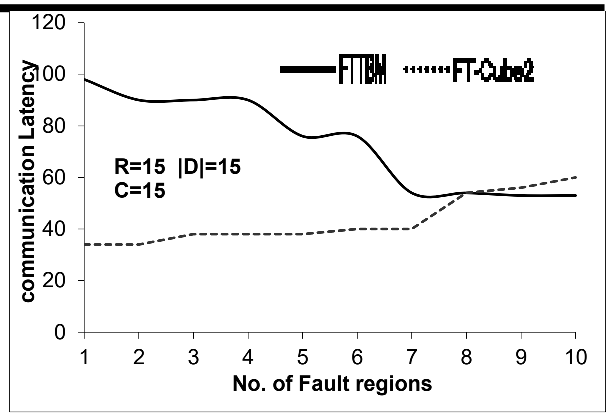


Figure 4 Communication Latency Vs. No. of fault regions

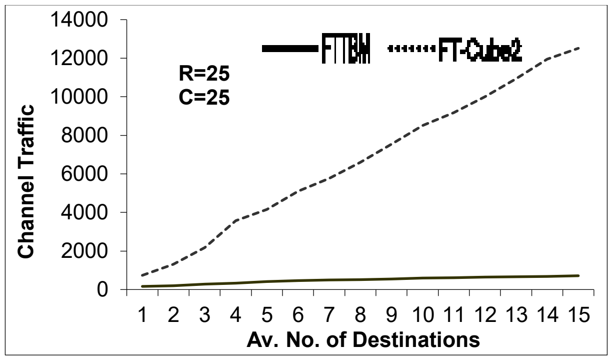


Figure 5 Channel Traffic Vs. Av. No. of Destinations

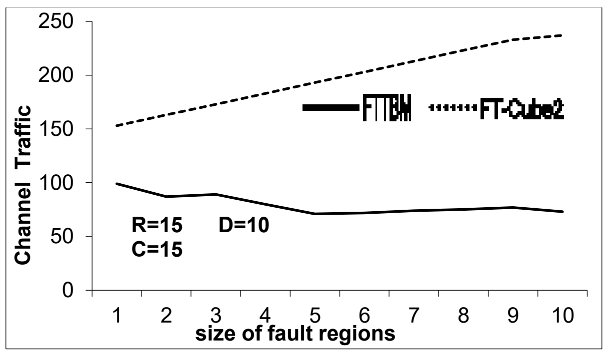


Figure 6 Channel Traffic Vs. Size of the fault region

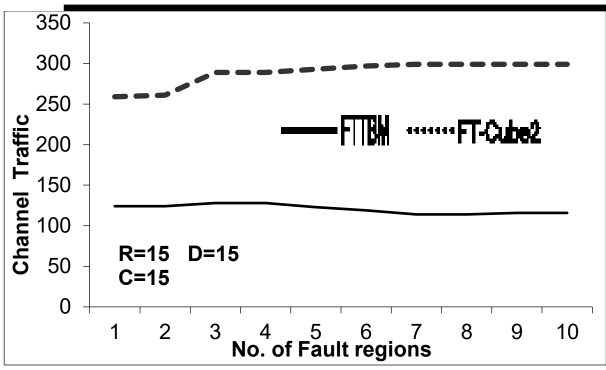


Figure 7 Channel Traffic Vs. No. of fault regions

Figure 5 plots the channel traffic for various values of average number of destinations, ranging from 1 to 20. The figures show that, the channel traffic computed by FTTBM algorithm is nearly constant as number of destinations increases. The channel traffic computed by FT-cube2 algorithm increases as number of destinations increases. The increasing rate of the channel traffic computed by FT-cube2 is large because each destination needs a separate message path.

Figure 6 plots the channel traffic for various sizes of one fault region, ranging from 1 to 100 and |D| equals 10. The figure shows that, the channel traffic computed by FTTBM algorithm is nearly constant as the size of the fault region increases, while the channel traffic computed by FT-cube2 algorithm increases. Also, this is because the path computed by FT-cube2 algorithm circles around the fault region.

Figure 7 plots the channel traffic for various values of number of fault regions, ranging from 1 to 10 and |D| equals 15. The figure shows that, the channel traffic computed by FTTBM algorithm is nearly constant as number of fault regions increases, while the channel traffic computed by FT-cube2 algorithm increases. This is because the number of fault regions which the path computed by FT-cube2 algorithm circles around them increases.

Clearly, in all tested cases, the channel traffic computed by FTTBM algorithm is less than that computed by FT-cube2 algorithm which is the mean target in this paper. Also, from the previous figures, the following notes can be observed:

- As size of the fault region and the number of fault regions increase, the communication latency computed by FTTBM algorithm decreases, while that computed by FT-cube2 algorithm increases. At small (large) fault region sizes and small (large) number of fault regions, the communication latency computed by FTTBM algorithm is larger (less) than that computed by FT-cube2 algorithm.
- As the number of destinations, size of the fault region, and number of fault regions increase, the channel traffic computed by FTTBM algorithm is nearly constant, while that computed by FT-cube2 algorithm increases.

5. Conclusion

In this paper, a new fault tolerant multicast algorithm, for wormhole routed 2D mesh multicomputers is introduced. The proposed algorithm is a tree-based multicast algorithm that works perfectly for the most commonly encountered faults in 2D mesh networks. It requires only local knowledge of all faults. It is effective even in the presence of a large number of faults and big fault size. The proposed algorithm is proved to be deadlock free. Channel traffic and communication latency are considered and calculated. Performance results of a simulation study on different 2D mesh networks are discussed to compare the proposed algorithm with a previous algorithm. The results show that the proposed algorithm is better and effective in most tested cases.

Acknowledgement: The author would like to show appreciation to Dr. Mohammed Abd-Elbaky at Fayoum University for providing help and support needed to conduct this research.

References

K. Hwang (1993). "Advanced Computer Architecture: Parallelism, Scalability, Programmability." McGraw-Hill, New York.

J. Duato (1993). "A New Theory of Deadlock-Free Adaptive Routing in Wormhole Networks." Proc. Symp. Parallel and Distributed Processing, pp. 64-71.

Rakesh Matam and Somanath Tripathy (2016). "Secure Multicast Routing algorithm for Wireless Mesh Networks." Hindawi Publishing Corporation, Journal of Computer Networks and Communications, Article ID 1563464, 11 Pages.

H.-H. Chang and G.-M. Chiu (2002). "An Improved Fault-Tolerant Routing Algorithm in Meshes with Convex Faults." parallel computing vol. 28, pp. 133-149.

R.V. Boppana and S. Chalasani (1995). "Fault-tolerant wormhole routing algorithms for mesh networks."IEEE Trans. Computers, 44(7):848–864.

S. Park, JH. Youn, B. Bose (2000). "Fault-tolerant wormhole routing algorithms in meshes in the presence of concave faults." in Proc. of International Parallel and Distributed Processing Symposium (IPDPS), p. 633-8.

Masoud E. Shaheen, Ahmed Abukmail (2012). "A Fault Tolerant Deadlock-Free Multicast Algorithm for 2D Mesh Multicomputers." The Journal of management and Engineering Integration (JMEI), Vol. 5, No. 2.

D. Xiang, Y. L. Zhang, and Y. Pan (2009). "Practical Deadlock-Free Fault-Tolerant Routing in Meshes Based on the Planar Network Fault Model. " IEEE Transactions on Computers Vol. 58, No. 5, 620-633.

Dong Xiang (2011). "Deadlock-Free Adaptive Routing in Meshes with Fault-Tolerance Ability Based on Channel Overlapping. " IEEE transactions on dependable and secure computing, vol. 8, no. 1.

Koffka Khan and Wayne Goodridge (2015). "Fault Tolerant Multi-Criteria Multi-Path Routing in Wireless Sensor Networks." I.J. Intelligent Systems and Applications, 06, 55-63.

T. Pfeiffenberger, J. Du, P. Arruda, A. Anzaloni, (2015)." Reliable and Flexible Communications for Power Systems: Fault-tolerant Multicast with SDN/OpenFlow", 7th IFIP on New Technologies, Mobility and Security (NTMS), 27-29.

X. Lin and L.M. Ni (1991). "Deadlock-Free Multicast Wormhole routing in Multicomputer Networks." In Proc. of the Intl. Symp. On Computer Architecture, pp. 116-124.

P. Mohapatra (1998). "Wormhole Routing Techniques for Directly Connected Multicomputer Systems." Dept. of Electrical and Computer Engineering, Iowa State University.

X. Lin, P. K. McKinley, and L. M. Ni (1994) "Deadlock-Free Multicast Wormhole Routing in 2-D Mesh Multicomputers." IEEE Trans. on Parallel and Distributed Systems, vol. 5, no. 8.

Integrating a Business Minor with an Engineering Course of Study

Tony Lewis[1]
Tomas Oppenheim[2]
[1]*California State University – Maritime*
[2]*California State University - Maritime*
tlewis@csum.edu; toppenheim@csum.edu

Abstract

Students who completed an undergraduate degree in engineering with a cross-disciplinary business focus believe themselves to be more effective employees than their peers who did not receive a cross-disciplinary education. There also appears to be strong industry demand for engineering graduates with combined training in business and weakening demand for graduates with a more focused engineering education. Yet most engineering programs do not seem to be responding to these facts and trends by integrating engineering and business programs of study. In this paper, the authors assess the incentives and disincentives that may exist around incorporating a business minor into an engineering course of study. The authors employ survey data and secondary research to gain insight into the level of student awareness and demand for this type of interdisciplinary training. Findings of this study suggest that the potential for collaboration between engineering and business departments is high. The authors explore some reasons why collaboration has often failed to materialize and document some reasons why it may benefit all major stakeholders to overcome those barriers.

1. Introduction

While many universities have both business and engineering departments, there is often little collaboration between these academic units (Miller & Euchner, 2014; Waters, 2010). This barrier exists despite the numerous and significant opportunities for synergy with the potential to provide meaningful benefits to students, academic departments, and universities. In this paper, we assess the incentives and disincentives that may exist around incorporating a business minor into an engineering course of study.

The findings of this study are based on a survey of over fifty undergraduates pursuing an engineering degree and secondary research on the costs and benefits of integrating business and engineering courses of study. Ultimately, our study provides support for four main findings. First, our research suggests that the opportunities for educational synergy between business and engineering departments are strong and largely unexploited. Second, our findings indicate that there is strong demand among engineering undergraduates to attain business knowledge, even though many are somewhat uninformed of the potential benefits. Third, we find that industry demand for graduates with a joint focus on business and engineering is strong and likely to get stronger (Waters, 2010) while demand for graduates with knowledge that is more focused on engineering alone is weakening (Clough, 2004; Kenney & Dossani, 2005). Fourth, our research suggests that the wider benefits to academic departments and universities of integrating business and engineering programs are substantial.

We begin our analysis by presenting the results of our survey of undergraduate students. We then discuss the costs and benefits of integrating engineering and business courses of study for students, academic departments, and universities. We close with a discussion of our conclusions, the limitations of this study, and potential avenues for future research.

2. Survey of engineering students

We received 54 responses to our survey of undergraduate students at Cal Maritime, a small California State University campus located in Vallejo, California. The campus is uniquely focused on preparing students for careers in the maritime industry, but the program itself is quite similar to other engineering programs. Many of the students responding to the survey were mechanical engineering majors and not particularly focused on acquiring skills related to the maritime industry. We queried students about their awareness of the existence of the business minor, their interest in pursuing the minor, their perceived feasibility of successfully completing the minor, the opportunity costs they thought might exist due to pursuing the minor, and the benefits they imagined might accrue from successfully completing the minor.

2.1 Awareness

We were surprised to find that slightly over 50% of respondents did not know that a business minor was offered. This is a strong indication of the lack of coordination that can exist between the business and engineering departments, a problem that we will touch on in more depth later in the paper. Even fewer respondents had received meaningful information about the potential benefits of a curriculum with a joint focus on engineering and business. Seventy-six percent (76%) of respondents received very little, or no information about the benefits of pursuing a business minor.

2.2 Interest

Slightly over half of the students surveyed were either slightly interested, or very interested in pursuing a business minor. We were surprised by this high number given that most respondents had no previous awareness of the existence of a business minor.

2.3 Feasibility

Sixty-five percent (65%) of respondents saw the prospect of completing a business minor as somewhat, or very feasible. The main concern of students centered around the heavy credit load already tied to the engineering major. Students already planning to take five years to complete their undergraduate degree were more likely to perceive successful completion of a business minor as feasible.

2.4 Opportunity costs

Twenty percent (20%) of respondents perceived the pursuit of a math minor as being more important than completing a business minor. We were surprised to find this response so low. For engineering majors at Cal Maritime, completing a math minor requires about half the courses required to complete the business minor. Moreover, with knowledge of the business minor and its potential benefits being so low, we expected the benefits of the math minor, a well-advertised offering for engineering students at Cal Maritime, to be perceived even more favorably when compared to the business minor. 14% of respondents indicated that they found opportunities related to internships to be more important than pursuing a business minor and thought that pursuit of a business minor might interfere. About 12% thought that maintaining an active social life was more important.

2.5 Perceived Benefits

Respondents perceived mainly career-oriented benefits relating to the completion of a business minor. Many students, about 24%, thought that pursuing a business minor would help them to move into a project management position. About 9% were interested in pursuing opportunities in entrepreneurship and perceived a business minor as being potentially helpful. Being a maritime university, a few respondents thought that the business minor would help them to eventually become a chief engineer of a ship.

3. Benefits to students

Of all the benefits associated with integrating business and engineering courses of study, most accrue directly to the students themselves. Students who completed an undergraduate degree in engineering with a cross-disciplinary business focus believe themselves to be better at teamwork, communication, project management, and general professionalism than their peers who did not receive a cross-disciplinary education (Ashton, et al., 2012). These results were found after students had spent some years in the workforce.

Business training also prepares students for the career trajectories that almost half of them are likely to pursue. Up to 50% of engineering graduates will pursue managerial or project management roles at some point in their career (Biddle, 1993). These students often own their own companies, perform consulting work, or manage product lines. Research suggests that, over time, career advancement causes engineers to become less interested in the narrow technical aspects of their jobs and more concerned with wider managerial issues (Biddle, 1993). Contemporary business trends are likely to continue to support increased demand for engineers with cross-disciplinary training. For example, one of the main principles of total quality management is to solve firm-wide problems through the implementation of cross-functional teams (McCahon& Lavelle, 1998).

Moreover, engineering careers that include a significant managerial component often pay substantially more than is common in a mainly technical engineering role. For example, a senior engineer at Boeing is compensated at around $150,000 annually (Salary.com). However, a project manager can get paid up to $250,000 (Salary.com).

3.1 Globalization concerns

In addition to the lure of higher salaries, globalization has pushed many U.S.-trained engineers away from highly focused technical engineering roles and into managerial roles. This is due to the fact that engineering firms can hire highly qualified technical engineers from overseas locations at a fraction of the cost when compared to U.S.-trained engineers (Clough, 2004; Kenney & Dossani, 2005). India's engineering schools produce about 260,000 highly qualified graduates per year who will work for salaries much lower than is common in the U.S. (Clough, 2004). The situation is similar in Russia and China (Clough, 2004).

At U.S. firms, most repetitive technical work is now outsourced to workers in developing nations due to lower costs of labor (Clough, 2004; Kenney & Dossani, 2005). When hiring for managerial or systems engineering roles, both requiring knowledge of business, U.S.-trained engineers who have a well-rounded, liberal education and who can think critically and analytically are still in high demand (Wulf, 1998). The ability to supply employees qualified to think critically and to come up with unique solutions to practical problems has been a traditional strength of the U.S. liberal education system (Zakaria, 2015).

3.2 Employment with startup firms

According to census bureau data from 1976-2005, over 65% of new jobs in the U.S. came from startup firms (Haltiwanger, Jarmin & Miranda, 2008). The US government has recognized the importance of integrating business and engineering training, investing millions of dollars in these efforts through the National Science Foundation which supports technology commercialization efforts, and through the Jumpstart Our Business Startups (JOBS) Act which supports startup firms focused on commercializing new technologies (Winkler, et al., 2015).

Engineers with an educational background that have a joint focus on business may be in a unique position to benefit from these government programs. Research suggests that people with this educational background have an increased ability to launch impactful startup firms focused on the commercialization of new technologies (e.g., Atkinson & Mayo, 2016; Duval-Couetil, 2013; Menzies & Paradi, 2001). Engineering program graduates who had some entrepreneurship training were more likely to launch startup firms and did so sooner after graduation than graduates who received no entrepreneurship training (Menzies & Paradi, 2001). Moreover, engineering majors who lack business training may delay efforts to commercialize promising innovations due to a perception that they require greater business expertise (Duval-Couetil, 2013).

3.3 Getting an employer to pay for graduate school

Some engineering students may be wary of dedicating time as undergraduates to the pursuit of a business minor because they think that they can simply go back to school for additional business training if opportunities to move into a management position materialize. However, this strategy may have significant weaknesses, particularly for those who would like to receive employer support for such training. Getting a business minor may allow engineering majors a more smooth transition into higher paying managerial positions by allowing them to more easily tiptoe into positions with increasing managerial responsibility. Relative to an engineering graduate with no business training, having a business minor will certainly increase the probability of moving into a position with some management responsibility. Having already demonstrated competence in a low-level management position may increase the probability of having an employer pay for a master of business administration degree (MBA) or other graduate training which is often required for higher level, better paying management positions.

In addition to being attractive candidates for an employer-sponsored MBA due to the increased probability of having already acquired some low-level management experience, engineering majors with a business minor can complete an MBA at much lower cost to the employer relative to someone with no business training. Generally, a business minor will fulfill all of the undergraduate prerequisite courses required to begin work on an MBA. So employees with a business minor already completed will be able to start graduate work right away. Employees without a business minor will generally require a semester or two to complete undergraduate prerequisite courses before they can begin graduate work. Hence, the cost to the employer is significantly lower for sponsoring an employee already holding a business minor to pursue an MBA compared to an employee who has not already completed a business minor.

4. Costs to students

A quick review of business minor information posted online by a range of universities that offer engineering undergraduate degrees indicates that students who pursue a business minor generally are required to enroll in about six additional courses. While there can be additional costs associated with taking these courses, the cost may be minimal given that full-time tuition often grants students a baseline number of courses they can enroll in, above which courses essentially become very low cost, or even free.

Given the high credit load that is characteristic of many engineering programs, pursuing a business minor will often require students to take an additional semester. With lack of planning or a late commitment to the minor, it may require an additional year since all courses required for the minor degree may not be offered in a single semester. With planning, or when enrolled in engineering programs that already require a minor course of study, a business minor may be completed with little-to-no additional time commitment from the student than what was already required as part of their engineering degree. Moreover, given the heavy demands of engineering programs, many engineering students will take more than four years to graduate regardless of their choice of minor. As indicated in our survey, such students may find it more feasible to complete a business minor with no additional time commitment than what they were already required to invest to complete their engineering degree.

4.1 Opportunity costs

Of course, there are opportunity costs associated with pursuing a business minor. Some students in our survey indicated that they would rather sharpen their core engineering skills by pursuing a more closely related minor course of study, like a math minor. However, there is considerable evidence in the literature that this may be a flawed strategy (e.g., Barnett, 2000; Datta, 2013; Svensson & Wihlborg, 2010). Barnett (2000) argues that we are living in “an age of supercomplexity” and that in this new age, students will benefit most from a broad, liberal education that emphasizes the ability to think critically and analytically to understand practical problems. In addition to making students more informed and engaged citizens (Svensson & Wihlborg, 2010), a more liberal education may help students to better navigate contemporary employment demands that require understanding a problem from different angles and the use of logic and intuition to come up with unique solutions to practical problems (Datta, 2013). The modern workplace requires employees who are aware and tolerant of international differences, which a more broad educational background can also help to provide (Svensson & Wihlborg, 2010). In response to these demands, the Association of American Colleges and Universities has begun the "Liberal Education and America's Promise Initiative," (LEAP). The LEAP website states that:

"*Through LEAP, hundreds of campuses are making far-reaching educational changes to help all their students—whatever their chosen field of study—acquire the broad knowledge, higher order capacities, and real world experience they need to thrive both in the economy and in a globally engaged democracy* (About LEAP, n.d.)"

Students who emphasize career-focused or instrumental learning may be limiting their own growth potential by investing in an overly-specialized education. While such an approach may be helpful for performing well in one's first job out of college, it is unlikely to serve students well as they move into the second and third job along their career trajectories (Datta, 2013). What today's engineering students will need to succeed in a long-term, evolving career path that characterizes the modern work environment is the ability to think from different perspectives and the ability to learn new skills that comes with a broader, more liberal education background (Datta, 2013).

While there is an opportunity cost associated with pursuing educational opportunities less closely related to one's major, students who do so are generally able to gain a more complete understanding of the problems they will be confronted with in the workplace (Barnett, 2000; Datta, 2013). They have a deeper understanding of the diverse world they live in and are better able to navigate the ambiguous and varied environments that characterize the modern workplace (Svensson & Wihlborg, 2010). This broader understanding will likely result in students who are more effective employees and better primed for advancement than students with a narrower academic focus (Barnett, 2000; Datta, 2013).

5. Benefits to departments & universities

Academic departments and universities may also benefit by integrating business and engineering courses of study. Most obviously, they may benefit from increasing enrollment. However, since the benefits for engineering majors to pursue some business courses may be higher than for business majors to pursue some engineering courses (Waters, 2010), the increased enrollment may be somewhat lopsided in favor of business departments. Engineering departments may generate increased enrollment themselves by offering new majors that rely on interdisciplinary cooperation, like a degree in engineering management. Industry demand for graduates with training in engineering management is strong and likely to get stronger (Waters, 2010). Business departments are also likely to experience greater enrollment in masters programs as engineering program graduates with business training move into positions with increasing management responsibilities (Waters, 2010).

As discussed in the Benefits to Students section, there is also high industry demand for engineering students with training in entrepreneurship. However, in 2010, only 23 out of the roughly 600 engineering schools in the US offered formal entrepreneurship training (Waters, 2010). Brown and MIT are the only engineering programs to offer a full major in entrepreneurship (Waters, 2010). Given the high industry demand for this type of training (Haltiwanger, Jarmin & Miranda, 2008) and the relatively small number of academic programs offered (Waters, 2010), this may be an area that is particularly attractive for universities to consider expanding into. Engineering departments that are early movers into engineering management or entrepreneurship training may be able to increase enrollment and gain industry support for academic programs that are highly responsive to business demands.

Business and engineering departments may both benefit from interdisciplinary cooperation through the production of insightful, unique research. Given the historic lack of integration between business and engineering departments (Miller & Euchner, 2014; Waters, 2010), scholarship that blends engineering and business disciplines may be a particularly fruitful area of research for faculty members to pursue. Increased collaboration between departments may also open up opportunities for joint teaching of classes that may benefit both students and faculty. Moreover, collaboration with business professionals may help to generate revenue-producing patents that can benefit departments, universities, and businesses.

Lastly, research suggests that many engineering students may not be highly engaged in their education due to reliance on rote learning and a highly technical or specialized focus (Miller & Euchner, J. 2014). Making the program of study more applied in nature may make engineering majors more attractive to more students and may help to increase diversity, a primary goal of many engineering programs (Miller & Euchner, J. 2014). In particular, research suggests that women may respond more strongly to a purpose-based career that makes a meaningful, observable difference in other people's lives (Miller & Euchner, J. 2014). Business training may allow students a more nuanced understanding of the wider sociological needs that the work of engineers responds to and greater knowledge of the broader impact of their efforts on society. This may lead to a more diverse student body, a primary goal of many universities and engineering departments (Miller & Euchner, J. 2014).

6. Costs to departments & universities

Compared to the possible benefits, the potential costs of integrating an engineering and business course of study may be quite small, especially for universities with existing engineering and business departments. Yet some costs are inevitable. Of course, with increased enrollment can come scheduling challenges. Courses may also have to be changed slightly; different examples used, or slight shifts in focus to be relevant to students with different career goals than the traditional business student.

As we have found in our survey of engineering students, it is critical that the business minor be promoted to students so that they are aware of its potential benefits. There may be some costs associated with attracting students. However, these costs may be quite minor given the high level of interest that students in our sample described, even with minimal promotion efforts made previously. Our survey suggests that an emphasis on career relevance may be the most effective means of promotion.

We anticipate that the biggest challenge for universities may be coordinating a new program of study between business and engineering professors who may have very different research focuses and professional norms and jargon. Faculty may see little personal benefit in integrating with another department when their own research may often have a very narrow focus. Faculty may also worry about a departmental shift to an area of scholarship where their own skill development may be lacking. We hope that the opportunities for impactful research and the wider benefits to students and universities will encourage faculty members to overcome these barriers.

7. Conclusions, limitations & future research

Our research suggests that the opportunities for educational and research synergy between business and engineering departments are strong and largely unexploited. We identify three main reasons for the historic lack of coordination. First, with resource scarcity becoming an increasing concern at U.S. universities, department administrators of both engineering and business departments may be wary of committing resources to curriculum development that is outside of what are historically considered core courses (Waters, 2010). Second, the dominance of business schools at many universities (Waters, 2010) may create a disincentive for engineering departments to collaborate out of concern that they may lose control over their own curriculum. Third, because the synergy is stronger for an undergraduate engineering major to take some business training than for an undergraduate business major to take some engineering courses (Waters, 2010), business departments may disproportionally benefit from increased enrollment. This may further decrease the incentives for engineering departments to initiate interdepartmental collaboration.

Despite these challenges, we find that there is strong demand among engineering undergraduates to attain business management skills. Moreover, we find that demand in industry for engineering graduates with cross-disciplinary business training is strong and getting stronger (Waters, 2010). At the same time, globalization is weakening industry demand for U.S.-trained engineering graduates with a highly focused skillset (Clough, 2004; Kenney & Dossani, 2005). Ultimately, we find that the benefits of interdepartmental collaboration between engineering and business departments for all major stakeholders (students, academic departments, and universities) are numerous and significant. Indeed, given the fact that students with a joint engineering and business focus are more likely to commercialize technologies and to create high-paying jobs by founding promising new startup firms (Atkinson & Mayo, 2016; Duval-Couetil, 2013; Menzies & Paradi, 2001), the benefits of this type of cross-disciplinary education may even extend more broadly to entire nations and economies.

There are some limitations to our study. First, because the career benefits are often greater for an engineering major to pursue some business courses than for an undergraduate student primarily focused on business to take some engineering classes (Waters, 2010), we focused our analysis on the former type of cross-disciplinary study. Future researchers may benefit from assessing the advantages that may be gained by business undergraduate majors who pursue a limited range of engineering courses.

Also, while we focused our analysis on the perspectives of students currently enrolled as engineering majors, future researchers may benefit from expanding research on whether engineering students who have graduated think that pursuing a business education as an undergraduate benefitted them, or would have benefitted them had they pursued it. Expanding scholarship into these important areas may help university administrators and faculty members to better understand the costs and benefits associated with facilitating collaboration between business and engineering departments. Ultimately, we hope that this knowledge may spur changes to the curriculums of engineering and business programs that may help students to realize the substantial benefits that appear possible from this type of academic collaboration.

8. Survey questions & response totals

Were you aware before today that a business minor is offered at Cal Maritime?

-Yes (29)
-No (28)

How much information have you received about the benefits of a curriculum with a joint focus on engineering and business?

-Quite a lot. Enough to answer all of my questions (1)
-Some, but not enough to answer all of my questions (12)
-Very little. Most of my questions have not been answered (20)
-None (24)

How interested are you in pursuing a business minor?

-Very interested (6)
-Slightly interested (27)
-Slightly uninterested (11)
-Definitely not interested (13)

If you are interested in pursuing a business minor, please list a few reasons that this course of study interests you. Number the list, placing the number one next to the most important or influential reason, a two next to the second-most important/influential, and so on.

-Becoming more familiar with business to advance career path (23)
-Expanding education (20)
-Want to own a small business (9)
-Interested in project management (10)

How likely do you think you are to successfully complete the business minor program if you choose to pursue it?

- Very likely (13)
-Somewhat likely (22)
-Not very likely (10)
-Highly unlikely (8)

If you are not interested in pursuing a business minor, please list a few reasons why you do not have interest in pursuing this course of study. Number the list, placing the number one next to the most important or influential reason, a two next to the second-most important/influential, and so on.

-Lack of interest (12)
-Course load is already too large (23)

Are there other opportunities outside of your regular engineering curriculum that you perceive as being more important? Please list these conflicting opportunities below. If not, simply write “no” below

-Math minor (11)
-Social life (7)
-Internships (8)
-Part-time work (1)

Are you aware of any careers that require a split business/engineering focus? If so, please list the careers below. If not, please simply write “no” below.

-Project management (12)
-Entrepreneurship (5)
-Chief engineer of a ship (2)

9. References

About LEAP (n.d.). Retrieved from https://www.aacu.org/leap

Ashton, M., Baily, J., Coomber, N., Goodell, K. & Weiland, K. (2012). The value of multidisciplinary engineering education after graduation: An empirical study.

Proceedings of the 2012 Industrial and Systems Engineering Research Conference

Atkinson, R. & Mayo, M. (2010). Refueling the U.S. innovation economy: Fresh approaches to science, technology, engineering and mathematics (STEM) education. The information Technology and Innovation Foundation. 1: 1-178

Barnett, R. (2000). Realizing the university in an age of supercomplexity. Buckingham: Open Univeristy Press.

Biddle, J. and Roberts, K. (1993). Private Sector Scientists and Engineers and The Transition to Management. The Journal of Human Resources. 29: 82-107.

Clough, G. (2004). At the crossroads: The future of engineering education. *Journal of Design Management:* July, 2004. 42-47.

Datta, A. (2013). Broad education, not narrow specialization, is the need of the hour. Businessline.

Duval-Couetil, N. (2013). Assessing the impact of entrepreneurship education programs: Challenges and approaches. Journal of Small Business Management: 51, 394-409.

Haltiwanger, J. Jarmin, R. & Miranda, J. (2008). Business formation and dynamics by business age: Results from the new business dynamics statistics. Center for Economic Studies, US Census Bureau.

Kenney, M. and Dossani, R. (2005). Offshoring and the Future of US Engineering: An Overview. The Bridge. 35:5-12.

McCahon, C. & Lavelle, J. 1998. Implementation of cross-disciplinary teams of business and engineering students for quality improvement projects. *Journal of Education for Business.* 73: 150-157.

Menzies, T. & Paradi, J. (2001). Encouraging technology-based ventures: Entrepreneurship education and engineering graduates. New England Journal of Entrepreneurship. 57-64.

Miller, R. & Euchner, J. (2014). The future of engineering education. Research-Technology Management. January-February 2014.

Svensson, L. & Wihlborg, W (2010). Internationalising the content of higher education: The need for a curriculum perspective. Higher Education. 60: 595-613.

Waters, R. (2010). Time to think outside the box? Technical entrepreneurship and engineering management education. Engineering Management Journal: *21*

Winkler, C., Troudt, E., Schweikert, C., & Schulman, S. (2015). Infusing business and entrepreneurship education into a computer science curriculum ; A case study of the STEM virtual enterprise. Journal of Business and Entrepreneurship: Fall 2015.

Wulf, W. (1998). The Urgency of Engineering Education Reform. New Paradigm for Engineering Education. 28:28-30.

Zakaria, F. (2015). In Defense of a Liberal Education. W.W. Norton and Compan

Reservoir Eutrophication Management Using SA base Pareto Multi-Objective Optimization

Xiaojue Tao, Ph.D[1]
Wen Fang, Ph.D[2]
Shoou-Yuh, Chang, Samuel Massie Chair Professor[3]

[1] *North Carolina Agricultural and Technical State University, Greensboro, NC 27411*
[2] *Network Control Division, Shanghai Electric Apparatus Research Institute*
[3] *North Carolina Agricultural and Technical State University, Greensboro, NC 27411*

Corresponding E-mail: t_crystal2004@hotmail.com; fangw@seari.com.cn

Abstract

Water system eutrophication is a world-wide issue for sustainable development. The aging process of natural water systems has been greatly accelerated by human activities. The water systems contain a harmful overdose of nitrogen and phosphorus caused by fertilizer, human activities, unplanned urbanization, special weather patterns, and so on. Excessive nutrients in water systems accelerate the growth of algae and worsened their ecosystems. This research provides an optimization remediation planning to control the reservoir eutrophication level. The land use management problem in the reservoir watershed is formulated as a two-objective optimization model. Maximize the economic income and minimize the environmental issues are two conflicting objectives addressing reservoir eutrophication management. Then a general simulated annealing algorithm was extended to discover the Pareto front of the two-objective optimization problem and generate optimal land use plans for stakeholders to choose according to their preferences. This optimization method can recommend the best land use plans of a reservoir watershed to prevent eutrophication.

1. Introduction

The concept of eutrophication as the process in which water bodies grow more productive, and proposed a classification mechanism based on plant production and nutrient availability. After several decades, researchers discovered that “good nourishment” had considerable negative environmental impacts on fresh water systems (Vollenweider, 1968; Rodhe, 1969; Vollenweider and Dillon, 1974; Carlson, 1977). The water systems contain a harmful overdose of nitrogen and phosphorus caused by fertilizer, human activities, unplanned urbanization, special weather patterns, and so on. Excessive nutrients in water systems accelerate the growth of algae and worsened their ecosystems. Based on the eutrophication status, decision makers can choose different remediation plans for eutrophication. These remediation plans need to achieve objectives and expectations in different aspects, including economic benefits, environmental impacts, time consuming, society influence, and so on (Reckhow and Chapra, 1983). Since these objectives conflict, a multi-objective optimization method was developed to generate optimal plans that achieve compromises between two conflicting objectives.

2. Literature review

Multiple objectives are considered in many optimization applications, including management, business, engineering and others. However, it is not easy to optimize multiple objectives simultaneously because they often conflict with one another (Stadler, 1988; Deb, 2014). Conventionally, multiple objectives are converted either to a single aggregated objective or constraints and one objective. The converted single objective will be optimized to find a single optimal solution, ideally a global optimum, that satisfies the physical restrictions of the system or process and reaches a compromise among all objectives (Marler and Arora, 2004; Wang and Liang, 2004). Many previous studies used classic methods, such as weighted aggregation, goal programming (Tabucanon, 1988), and ε-constraint, to seek global optima. The art of compromise is the core of this process, which requires: (1) the objectives are convertible to constraints; (2) a priori knowledge about the relative importance of the objectives; (3) the aggregated function leads to only one solution; (4) trade-offs between objectives cannot be easily evaluated. Moreover, the solution may not be attainable because the solver may be trapped into a local optimum.

It is more challenging to solve a multi-objective optimization problem because it does not seek a single solution, but rather a set of acceptable alternative optimal solutions. In this research, Simulated annealing (SA) is selected as the optimization algorithm. SA is a global search algorithm which is able to escape from a local minimum by applying the hill-climbing strategy. SA mimics the metaphor of the process of physical annealing with solids, in which a crystalline solid is heated and then allowed to cool very slowly until it achieves its most regular possible crystal lattice configuration (i.e., its minimum lattice energy state); thus, is free of crystal defects (Kirkpatrick, 1984). If the cooling schedule is sufficiently slow, the final configuration results in a solid with superior structural integrity. Simulated annealing establishes the connection between this type of thermodynamic behavior and the search for global minima for a discrete optimization problem (Lundy and Mees, 1986). Furthermore, it provides an algorithmic method for exploiting such a connection (Rossier et al., 1986).

3. Research objective

The objective of this research is to develop a multi-objective optimization algorithm for optimal land uses. This method is capable of generating Pareto optimal development plans in the watershed by considering economic benefits and environmental impacts. The Pareto optimal development plans will provide the best options for decision makers to choose a proper plan based on their preferences.

4. Methodology

The Pareto front provides a set of Pareto optimal solutions for multi-objective problems. This research modified the simulated annealing algorithm to find the optimal solution set. The solutions are more visual and flexible.

4.1. Multi-objective optimization

For generating a set of acceptable optimal solutions, the Pareto optimality concept is introduced. In fact, the multi-objective optimization aims to discover all Pareto optimal solutions (Abbass et al., 2001). Formally, a Pareto optimal solution $\mathbf{x}^* = [x_1^* \; x_2^* \; x_3^* \cdots]$ is recognized to Pareto-dominate other solutions $\mathbf{x} = [x_1 \; x_2 \; x_3 \; \ldots]$, in a minimum context, if and only if (Coello et al., 2002)

$\forall i \in \{1,\ldots,N\}, f_i(\mathbf{x}^*) \le f_i(\mathbf{x})$, and
$\exists j \in \{1,\ldots,N\}, f_j(\mathbf{x}^*) < f_j(\mathbf{x})$.

In other words, a solution $\mathbf{x}^*$ is called to be the Pareto optimal solution if and only if no other solutions dominate it (Ngatchou et al., 2005). All Pareto optimal solutions form the Pareto front (Van Veldhuizen and Lamont, 1998). Figure 1 illustrates the Pareto fronts for a two-objective optimization problem. If the optimization problem seeks the maximum of objective 1 and the maximum of objective 2, the upper right edge of the ellipsoid is the Pareto front, and all other solutions in the ellipsoid are dominated by the Pareto front. Similarly, if seeking the maximum of objective 1 and the minimum of objective 2, the lower right edge of ellipsoid is the Pareto front.

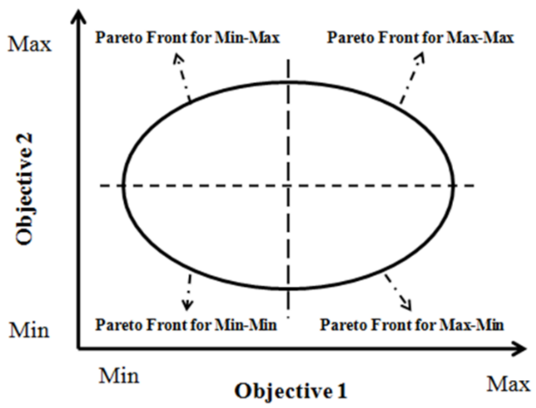


Figure 1. Illustration of Pareto front of two-objective optimization

The Pareto front provides a set of Pareto optimal solutions to stakeholders for decision making, and these solutions guarantee that there are no other solutions that could improve a single objective without compromising the other objectives (Podinovskii and Nogin, 1982; Kaufman and Rousseeuw, 2008). The stakeholders obtain the useful information from the Pareto front, and every party on the table could set up their bargaining strategies with the knowledge of the other parties' choices. If parties are cooperative, their initial goals are downgraded during the bargaining process, and an optimal state exists along the Pareto front, see Figure 2 Obviously, according to different scenarios, the optimal state is possible to move along the Pareto front, which offers a great support to the stakeholders for decision making with or without prior expertise.

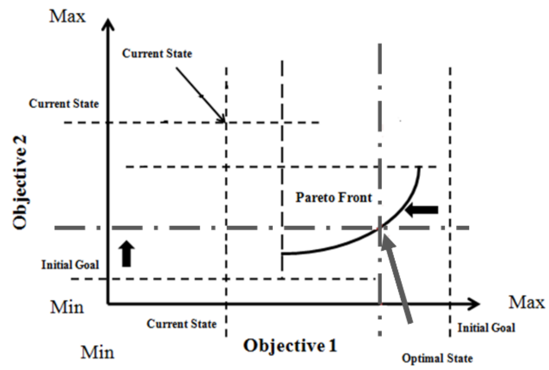


Figure 2. Illustration of the bargain process for a max-min two-objective optimization

4.2 Simulated annealing (SA)

With the development of the evolutionary algorithms and computational machines, their convergence speed and the other optimization performance have been greatly enhanced. These techniques are more appropriate for generic optimization problems, particularly when preferences are not available in advance. One heuristic evolutionary algorithms, SA, is well known for its hill-climbing strategy (Hwang, 1988; Johnson and Jacobson, 2002). This advantage allows SA to avoid being trapped in a local minimum by accepting a worse solution with a certain probability (Aarts and Van Laarhoven, 1985).

4.3. Modified simulated annealing for multi-objective optimization

According to the definition of the Pareto front mentioned in section 4.1, the difference between Pareto optimal solutions and other solutions is that a Pareto optimal solution is not dominated by any other solution, and this characteristic of a two objective min-min optimization is displayed in Figure 3. The larger dots are non-dominated solutions, and the other dots are dominated solutions. In other words, if a new solution **x** is non-dominated, it must satisfy one of the three following conditions.

Condition 1:

If there exists any $\mathbf{x}' \in X_{ND}$ satisfying $f_1(\mathbf{x}) < f_1(\mathbf{x}')$ and $f_2(\mathbf{x}) < f_2(\mathbf{x}')$, then $\mathbf{x} \in X_{ND}$.

Condition 2:

If $f_1(\mathbf{x}) < \min_{\forall \mathbf{x} \in X_{ND}} f_1(\mathbf{x}')$ and $f_2(\mathbf{x}) > \max_{\forall \mathbf{x} \in X_{ND}} f_2(\mathbf{x}')$, then $\mathbf{x} \in X_{ND}$.

Condition 3:

If $f_1(\mathbf{x}) > \max_{\forall \mathbf{x} \in X_{ND}} f_1(\mathbf{x}')$ and $f_2(\mathbf{x}) < \min_{\forall \mathbf{x} \in X_{ND}} f_2(\mathbf{x}')$, then $\mathbf{x} \in X_{ND}$.

In these conditions, X_{ND} denotes the set of solutions that are not dominated by any solution already exploited in the SA searching process, $f_1(\bullet)$ is the first objective function, and $f_2(\bullet)$ is the second objective function. According to the definition of the non-dominance, these three conditions could be adjusted to fit other scenarios of multi-objective optimization problems.

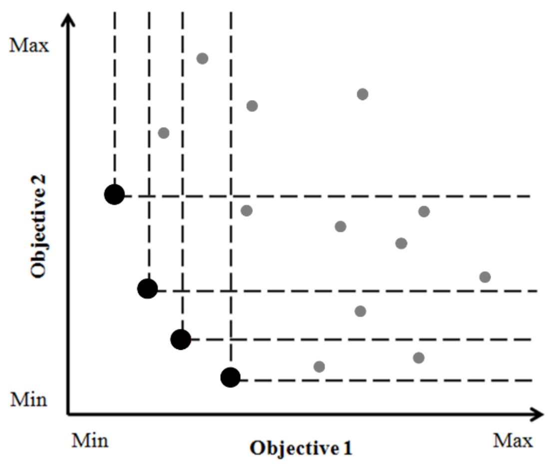


Figure 3. Pareto front for a two-objective min-min optimization problem

The main idea behind the modification is seeking the Pareto front along the evolutionary optimization process of either one of the objectives (Yao et al., 1999; Zhao et al., 2007). A modified decision is added after the performance comparison step, which results in updating the non-dominated solution set based on the three conditions for non-dominated solutions. The final non-dominated solution set constructs the Pareto front after the optimization process stops (Honkela, 1997; Kohonen, 1990; Michalewicz, 2013).

5. Experimental results

The Te-Chi Reservoir is a study case which is located in the upstream of Ta-Chia Creek in central Taiwan. It is the fourth largest (in terms of storage volume) reservoir in Taiwan with a maximum water surface area of 454 ha and initial design storage volume of about 2.32×10^8 m^3. The annual inflow is about 1.2×10^9 m^3, about five times the reservoir volume, but over three-fourth comes during the wet season. The watershed area is 592 km^2.

5.1. Multi-objective model

The economic development and environment protection always conflict with each other, which is also true in the reservoir watershed land use management. In this case study, two objectives are considered in the land use management of the Te-Chi reservoir watershed.

Objective 1: Maximizing the incomes from different land uses

$$Obj_{income} = \max\left(\sum_{i=1}^{n} \alpha_i \cdot x_i\right),$$

where α_i is the income of the i^{th} type of land use or development per hectare, x_i is the area of the i^{th} type of land use or development (hectare), and n is the total number of land use types.

Objective 2: Minimizing the impact on the environment (i.e., minimizing the TP concentration)

$$Obj_{environment} = \min\left(\sum_{i=1}^{n} \beta_i \cdot x_i\right),$$

where β_i is the TP export coefficient of the i^{th} type of land use or development per hectare (kg/ha · year).

Several constraints should be considered in the reservoir watershed land use management. These constraints are discussed as follows.

Constraint 1:

$$\mathrm{MTSI}(C) \le 40,$$

which indicates the trophic state is oligotrophic.

Constraint 2:

$$\sum_{i=1}^{n} x_i = 60252.67 \text{ ha},$$

which limits the total land availability for different uses in the watershed.

Constraint 3:

$x_i \ge 50$ ha, for i = 3 (residential use),

which guarantees the minimum land availability for residential development plan in order to preserve the native culture.

Constraint 4:

$x_i \ge 50729.52$ ha, for i = 4 (forest),

which ensures that the forest area is not allowed to be reduced any more.

Constraint 5:

$x_i \ge 285.6$ ha, for i = 1 (waterbody),

which enforces the minimum prevention area, including the water body and the river bank.

Non-negative constraints:

$$\forall x_i \ge 0,$$

which force all types of area should not be negative.

5.2. Application and modification

Figure 4 illustrates the temperature schedule of implementing the modified SA algorithm on the income objective. Since the income objective is to be maximized, the figure shows the increasing trend of income with one exception. Figure 4 shows that the second last accepted solution is a worsen one compared to the third last solution. However, it is the advantage of SA, which accepts a worsen solution for avoiding tapped in a local maximum, called a hill-climbing feature. Meanwhile, the set of non-dominated optimal solutions is exploited in the evolutionary optimization process with a focus on the income. The dots in Figure 5 show all non-dominated optimal solutions exploited, which form the Pareto front.

The Pareto front is a base-curve for the stakeholders to make their decisions responding to different aims. By applying the concept of the game theory, the results are more visual and flexible. Two stakeholders representing the sides of economic development and environment protection are setting initial goals and bargaining strategies based on the information extracted from the Pareto front. The best and worst conditions are shown in Table 1. The best status of the income is 607,260.2×10^3 USD, which means the stakeholder who only concerns the economic objective achieve best economic benefits, and must accept the worst of TP export of 40,857.6 kg. On the other side, when the environment objective reaches the best TP export of 16,306.1 kg, i.e., the minimum yearly TP export, the economic income is the lowest.

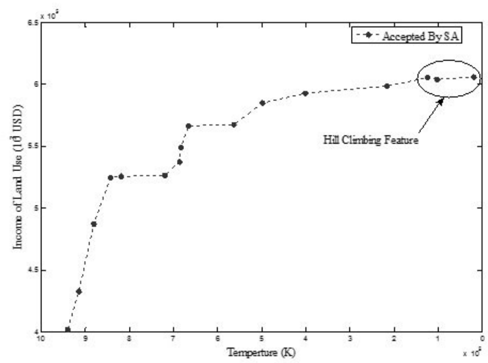


Figure 4. Cooling schedule of SA in objective 1 value

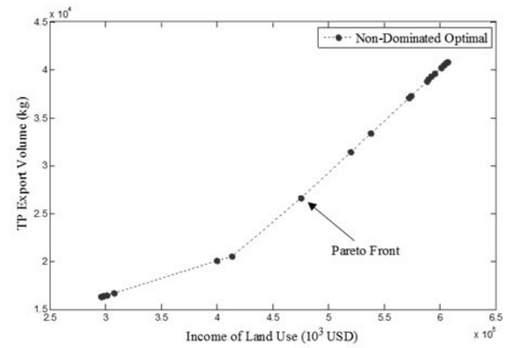


Figure 5. The Pareto front constructed using the modified SA

Table 1. The best and Worst Conditions Extracted from the Pareto Front

	Economic Objective (10^3USD)	Environment Objective (kg)
Best Status	607,260.2	16,306.1
Worst Status	295,938.7	40,857.6

The initial goal of the economic objective is 607,260,200 USD, and the initial goal of the environment objective is TP export of 16,306.1 kg per year. Assuming that each stakeholder would sacrifice 5% of the initial goal in every round of the bargaining process, the bargaining process shown in Figure 6 shows the bargaining steps and the Pareto front. The figure demonstrates that the bargaining points enter the dominated area after the ninth bargain, which means these bargaining points are not optimal because they are dominated by a non-dominated optimal solution on the Pareto front. It also implies that there always exists a non-dominated optimal when the bargaining process moves towards to the upper left side of the Pareto front in any way. This is greatly useful because the stakeholders are rarely cooperative with each other and the realistic bargaining process could be in any way.

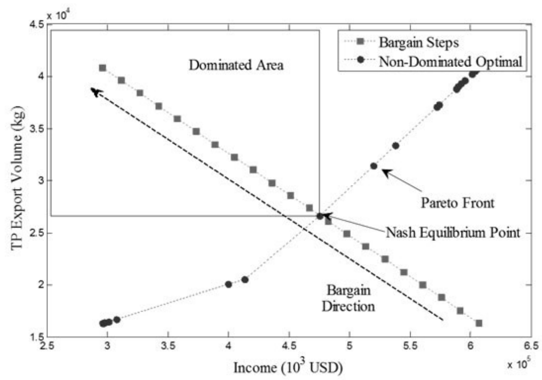


Figure 6. The bargaining process and Pareto optimal solution

This non-dominated optimal solution is the optimal solution of the Te-Chi reservoir watershed land use optimization problem without considering the hard constraint of TP export. The goal of the 9th bargaining step is (467,165.5, 27,353.3) and the optimal solution is (475,562.7, 26,632.8), which is known as the Nash equilibrium (Osborne and Rubinstein, 1994; Gibbons, 1997). In the Nash equilibrium, this non-dominated optimal solution is better than the bargain goal in terms of both the economic and environment objectives.

Table 2 lists this non-dominated optimal solution and the current land use in the Te-Chi reservoir watershed. The optimal solution reduces the size of farmland and grassland to very low levels, which means those two types of land use are not economically or environmentally efficient compared to the other types of land use. The optimal solution increases the income of land use as well as the TP export by 18.29% and 15.49%, respectively.

Table 2. The optimal solution of Te-Chi reservoir watershed land use optimization problem

	Current solutions	Optimal solutions	Difference
Obj1	402,030.0	475,562.7	73,532.7
Obj2	23,061.0	26,632.8	3,571.8
1	285.6	8,392.2	8,106.6
2	2,685.1	163.6	-2,521.5
3	2,065.7	2,673.0	607.3
4	50,729.5	50,749.4	19.9
5	4,185.6	10.2	-4,175.4

obj1: Economic Objective (10^3USD)
obj2: Environment Objective (kg)
1: Water Body (ha)
2: Farm Land (ha)
3: Residential Area (ha)
4: Forest (ha)
5: Grass Land (ha)

The increase in TP export is not optimistic in the final optimal solution, and it is necessary to check the water quality is eutrophic or not. If the maximum yearly TP assimilative capacity is set as 25,805.25 kg. Figure 7 demonstrates that when the bargain processes until the 9th round, it enters the Pareto front. The other two new Pareto optimal solutions reach the Nash equilibrium, which are listed in Table 3 and Table

4. These two solutions both decrease the TP export amount, and one of these two solutions also increases the annual income of land use by 11,716.4×10^3 USD. In both solutions, the residential area and grassland area should be greatly reduced. The grassland area is reduced because of the insufficient contribution of environmental protection, and the residential area is reduced to decrease the TP export, which results in the economic sacrifice.

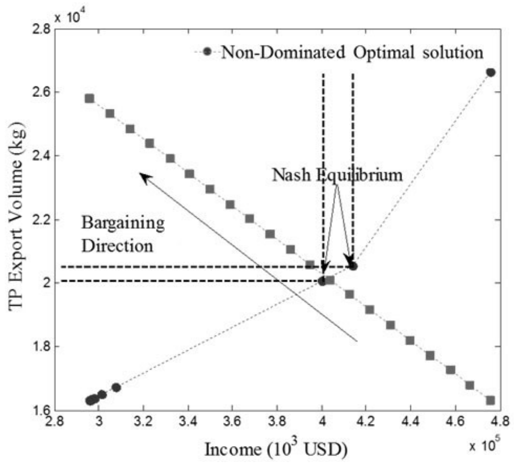


Figure 7. The bargaining process and Pareto optimal solutions under TP constrains

Table 3. The optimal solution 1 of the Te-Chi reservoir watershed land use optimization problem constrained by TP concentration

	Current land use	Optimal solution1	Difference
Obj1	402030.0	400206.0	-1823.9
Obj2	23061	20047.5	-3013.4
1	285.6	229.7	-55.8
2	2685.12	6615.1	3930.0
3	2065.6	52.2	-2013.4
4	50729.5	5288.4	2138.9
5	4185.6	201.4	-3984.2

Table 4. The optimal solution 2 of the Te-Chi reservoir watershed land use optimization problem constrained by TP concentration

	Current land use	Optimal solution2	Difference
Obj1	402030.0	413746.4	11716.4
Obj2	23061	20552.0	-2508.9
1	285.6	96.0	690.4
2	2685.12	8487.9	5802.8
3	2065.6	50.4	-2015.2
4	50729.5	50733.5	3.9
5	4185.6	4.7	-4180.9

6. Conclusion

The land use optimization problem in the reservoir watershed involves two conflicting objectives: maximizing the economic income and minimizing the environmental impacts. A multi-objective optimization method was developed to optimize these two objectives, and a general simulated annealing (SA) algorithm was extended to discover the Pareto front of the multi-objective optimization problem and generate optimal land use plans under different preferences. In the extended SA algorithm, along the cooling schedule, the non-dominated Pareto front is discovered while one of the objectives reaches its optimum. The initial land areas for different uses in the watershed can be extracted from the geographic information system (GIS) with the image processing technologies. When alternating the land areas among different uses, the accumulated TP concentration in the reservoir changes. Along with the cooling schedule, the Pareto front is obtained. The limitation of yearly allowable TP export volume is enforced by the constraint of the acceptable TP concentration by applying the zero-dimensional water quality model. For a given bargaining process, a Pareto optimal solution exists to dominate a possible agreement of the bargaining process, and yields to the Nash equilibrium, which means the maximum land use income without downgrading the water eutrophication status. Under different scenarios of the bargaining process, multiple Pareto optimal solutions can be found. Therefore, this approach can provide a set of optimal recommendations for the stakeholders to make proper decisions based on their preferences.

7. Reference

Abbass, H. A., Sarker, R., & Newton, C. (2001). PDE: A Pareto-frontier differential evolution approach for multi-objective optimization problems. In *Proceedings of the IEEE Congress on Evolutionary Computation,* (Vol. 2), pp. 971-978.

Aarts, E. H., & Van Laarhoven, P. J. (1985). Statistical cooling: A general approach to combinatorial optimization problems. *Philips J. Res.*, *40*(4), 193-226.

Carlson, R. E. (1977). A trophic state index for lakes. *Limnology and oceanography*, *22*(2), 361-369.

Deb, K. (2014). Multi-objective optimization. In *Search Methodologies*. Springer, pp. 403-449.

Hwang, C. R. (1988). Simulated annealing: Theory and applications. *Acta Applicandae Mathematicae*, *12*(1), 108-111.

Honkela, T. (1997). *Self-Organizing Maps in Natural Language Processing*. Doctoral dissertation, Helsinki University of Technology.

Johnson, A. W., & Jacobson, S. H. (2002). A class of convergent generalized hill climbing algorithms. *Applied Mathematics and Computation*, *125*(2), 359-373.

Kirkpatrick, S. (1984). Optimization by simulated annealing: Quantitative studies. *Journal of statistical physics*, *34*(5-6), 975-986.

Kaufman, L., & Rousseeuw, P. J. (2008). *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, Inc., Hoboken, NJ, USA.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464-1480.

Lundy, M., & Mees, A. (1986). Convergence of an annealing algorithm. *Mathematical programming*, *34*(1), 111-124.

Marler, R. T., & Arora, J. S. (2004). Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, *26*(6), 369-395.

Michalewicz, Z. (2013). *Genetic algorithms+ data structures= evolution programs*. Springer Science & Business Media.

Ngatchou, P., Zarei, A., & El-Sharkawi, M. A. (2005). Pareto multi objective optimization. In *Proceedings of the 13th IEEE International Conference on Intelligent Systems Application to Power Systems*, pp. 84-91.

Podinovskii, V. V., & Nogin, V. D. (1982). Pareto-optimal solutions of multi-criteria problems.

Rodhe, W. (1969). *Crystallization of eutrophication concepts in northern Europe. In: Rohlich, G.A. (Ed.), Eutrophication: Causes, Consequences, and Correctives.* Washington, DC: National Academy of Sciences.

Reckhow, K. H., & Chapra, S. C. (1983). *Engineering Approaches for Lake Management*. Boston, MA: Butterworth.

Rossier, Y., Troyon, M., & Liebling, T. M. (1986). Probabilistic exchange algorithms and Euclidean traveling salesman problems. *Operations-Research-Spektrum*, *8*(3), 151-164.

Stadler, W. (1988). *Multicriteria Optimization in Engineering & in the Sciences*, Plenum Press, New York.

Tabucanon, M. T. (1988). *Multiple Criteria Decision Making in Industry*. Elsevier Science Ltd.

Vollenweider, R. A. (1968). The scientific basis of lake and stream eutrophication, with particular reference to phosphorus and nitrogen as eutrophication factors. *Organisation for Economic Cooperation and Development.*

Vollenweider, R. A., & Dillon, P. J. (1974). *The Applications of the Phosphorus Loading Concept Eutrophication Research* (Vol. 13690). National Research council of Canada.

Van Veldhuizen, D. A., & Lamont, G. B. (1998). Evolutionary computation and convergence to a Pareto front. In *Proceedings of Late Breaking Papers at the Genetic Programming Conference,* pp. 221-228.

Wang, R. C., & Liang, T. F. (2004). Application of fuzzy multi-objective linear programming to aggregate production planning. *Computers & Industrial Engineering*, *46*(1), 17-41.

Yao, X., Liu, Y., & Lin, G. (1999). Evolutionary programming made faster. *IEEE Transactions on Evolutionary Computation*, *3*(2), 82-102.

Zhao, X., Gao, X. S., & Hu, Z. C. (2007). Evolutionary programming based on non-uniform mutation. *Applied Mathematics and Computation*, *192*(1), 1-11.

Using the Tools of Quality to Improve Production Operations at Wilmington Textile

J. S. Sutterfield
Tolulope R. Daramola
Florida A&M University
Tallahassee, FL
Pisces4914@earthlink.net

Abstract

Using the tools of quality in real-world production applications can reward a company with immense and readily quantifiable economic benefits. In the following paper a case is presented in which the tools of quality were applied to analyze an actual production problem in the textile industry. Using these tools it was possible to identify root causes of production problems down to the worker level, and to estimate the economic benefits of improving production operations as indicated by application of the tools. In this case, only the names were changed to protect the guilty.

1. Introduction

Wilmington Textile is one of the world's global leaders in the fabric industry. The Wilmington business base includes expensive fabrics for commercial and consumer applications. The company is headquartered in Wilmington, NC, and has several textile mills located throughout the U.S. and Canada. Wilmington employs approximately 50,000 people in these countries, but serves customers worldwide. Among Wilmington's numerous mills, is one located in Grayson, Kentucky. Production at the Grayson mill includes a wide variety of finished and semi-finished fine fabrics. Production quality at Grayson had always been above average, but during the past few months, the Grayson mill had begun to develop some problems in its production operations. Specifically, the Grayson Mill had begun to produce an excessive number of non-conforming Products (NCPs). Non-conforming products are those products that may have been over-produced, leftover from customer orders, or simply defective. NCP codes were defined by production management as follows:

- NC0 – Quality hold
- NC1 – Quality reject
- NC2 – Joblot produced against NCP joblot
- NC4 – Non-salable trim fill
- NC5 – Salvage product
- NC6 – Straight wind product
- NC9 – Cancelled order transferred to NC overrun
- NC10 – Production overrun transferred to NC overrun
- NC11 – Scheduled overrun transferred to NC overrun
- NC12 – Scheduled trim fill transferred to NC overrun
- NC13 – Manual transfer of inventory to NC overrun
- NC14 – Cancelled order or stranded bolt transferred to NC overrun

For efficient operations at the Grayson Mill, the total inventory (finished, unfinished and NCP) should be no more than 7,800 bolts with no more than 10% as NCP. (Note: A bolt of fabric is 100 yards.) Essentially, each NCP involves one bolt of fabric. In 2013, Grayson had held a below average total inventory of 6,000 bolts, but with NCP representing approximately 1,050 bolts. Thus, the NCP was 75% greater than that established for cost efficient operations. Moreover, this excess amount of NCP represented a working capital/inventory cost of $45,000 ($100/bolt). Even more alarming was the fact that the data from the first half of 2014 indicated a substantial increase in NCP inventory to 1,350 bolts. With an average of 7,500 bolts, this amount of NCP was 80% greater than that required for efficient mill operation at Grayson.

Always concerned about production costs, management at the Grayson mill decided to assemble a team of four knowledgeable employees to address the problem of non-conforming product. The team was assigned the task of reducing NCP at Grayson to not more than 10% of average inventory for 2015.

The team began by developing the following tentative action plan: 1) Survey and familiarization with available quality tools and processes; 2) Determine what data were available, and any additional data necessary for the analysis; 3) Collect all available data; 4) Perform measurements to acquire any data not presently available; 5) Analyze and process all the necessary information; 6) Draw conclusions as to root causes of NCP; 7) Recommend potential solutions to decrease NCP's to 10% or less; 8) Perform a limited implementation of preferred solution; 9) Monitor system to determine whether or not preferred solution corrects problem(s); 10) Develop remedial plan to implement preferred solution. 11) Implement remedial plan; 12) Monitor process(s) to ensure continuing compliance with correction(s).

2. Literature search and analysis

The team began its survey of the available quality improvement literature, by examining the work of William Edwards Deming. This research disclosed a number of approaches devised by Deming to manage change in order to improve quality. The first was the ***P-D-S-A cycle*** (Deming, 1986) as adapted from the original work of Dr. Walter A. Shewart (1939). Deming's P-D-S-A cycle can be seen below in **Figure 1**.

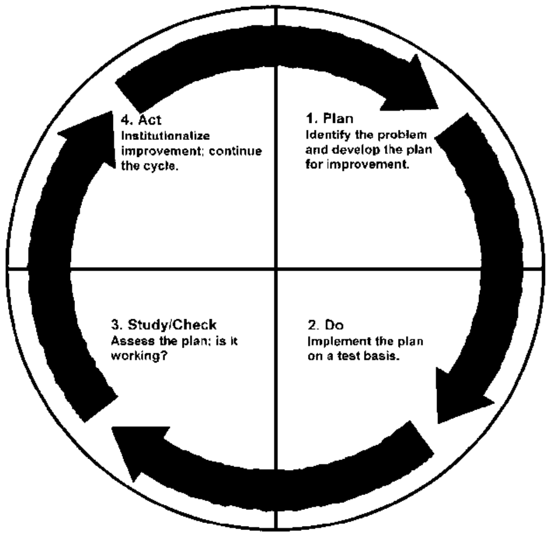


Figure 1: Deming's PDSA Wheel

The team was pleased to learn that their initial common-sense approach to the study was along the lines taught by Deming. The team's research further disclosed from Deming's famous 14 points that for a project to be successful it would need to have a champion (Deming, 1986). This requirement was further underscored in another work by Ahire (1997), which made the point that a lack of management leadership would preclude a successful quality program.

Proceeding, the quality team examined the ***Cause-and-Effect Diagram*** or ***Ishikawa Diagram***. This quality tool is named after its originator, Kaoru Ishikawa (Ishikawa, 1968, 1976) of Japan. Popularly it is known as the "Fishbone Diagram" due to its shape. It is used to allow graphical depiction of all possible causes of some production problem. The cause and effect diagram also may be used to identify and organize those factors needed to ensure success of some effort. Thus, it allows easy identification of the relationships among factors when studying processes, problem situations, or production planning. It is also used as an effective tool to identify causes of quality problems. When used in this application, it is known as ***Failure Mode Effects Analysis (FMEA)***. This diagram was adopted by Dr. Deming as a valuable device for improving quality. He used the diagram as one of his first tools in the quality management process. The Ishikawa Diagram is shown below in **Figure 2**.

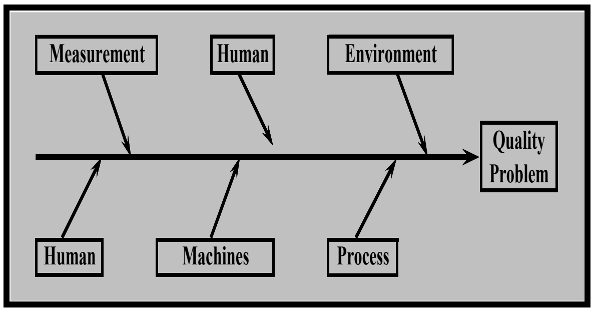


Figure 2: Ishikawa or cause & effect Diagram

As the Grayson quality team continued with their investigation of the tools for quality, they discovered a device for tabulating data called the ***Check Sheet*** or ***Defect Concentration Diagram***. The check sheet is considered to be one of the seven basic tools of quality (Tague, 2004), and provides a convenient means of recording data to facilitate later analysis. The check sheet is used to document information as to the frequency of some occurrence(s) of interest in an efficient, tabular format. This may be accomplished with a simple listing of occurrences, such as flaws or defects or failures, along with the number of times that each flaw occurs. Although the check sheet is quite useful when used in its simplest form, its utility may be significantly enhanced by devising a form specific to the system under analysis. An example of a Check Sheet is shown in **Figure 3**.

Product: Flywheel Housing | **Date:** 6-17-04

Manufacturing Stage: Final Inspection | **Location:** Wentzville, MO.

Defect Type: Scar, Incomplete, Mis-shapen | **Inspector:** Addington

Total Number Inspected: 3,160 | **Lot No.:** 14,756

Remarks: All items inspected | **Order No.:** 371

Defect Type	Frequency of Occurrence	Sub-total
Surface Scars	///// ///// ///// ///// ///// ///// //	32
Cracks	///// ///// ///// ///// ///	23
Incomplete	///// ///// ///// ///// ///// ///// ///// ///// ///// ///	48
Mis-shapen	////	4
Miscellaneous	///// ///	8
	Grand Total	115
Total Rejects	///// ///// ///// ///// ///// ///// ///// ///// ///// ///// ///// ///// ///// ///// ///// ///// ///// /	86

Figure 3: Check Sheet

Along with defect identification, the defect concentration diagram is also useful for other purposes such as defect location and defect cause (Ishikawa, 1986).

As the Grayson quality team continued with its investigation , they found that the check sheet is often used in conjunction with the ***Histogram***. The *histogram* was devised by the statistician Karl Pearson more than a century ago (Pearson, 1895), and evolved to meet the need for evaluating data that occur with various frequencies. The histogram makes this possible by allowing for a concise display of information in a bar graph format. The histogram is a powerful analytical tool in that it clearly and concisely portrays information as to the location, spread, and shape of the distribution of the data describing the functioning of the physical process. It also enables the analyst to perceive subtleties regarding the functioning of the physical process. It can also help to suggest both the nature of, and possible improvements to, the physical mechanisms at work in the process. An example of a histogram is shown in **Figure 4**.

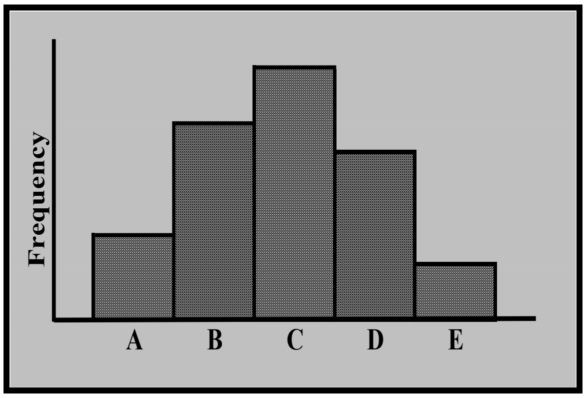


Figure 4: Histogram

The discovery of the histogram also occasioned the discovery of another closely related tool of quality known as the ***Pareto Chart***. The Pareto chart graphically displays the Pareto principle, a principle first enunciated by the Italian economist Vilifredo Pareto in 1906 (Bousquet and Busino, 1964). The Pareto principle states that 80% of the wealth of a country is concentrated in 20% of the population, and vice-versa. In effect, the Pareto diagram plots the information from check sheets as histograms in which the problem frequencies are plotted in the order of greatest to least. Pareto's insight as to the 80-20 rule has also been found useful in analyzing production defects. Modified for application to the production context, the rule would state that about 20% of product defects occur about 80% of the time, and about 80% of the product defects occur about 20% of the time. Pareto charts are extremely useful because they can be used to identify those problems having the most cumulative effect on a system, and thus screen out the less significant factors. Ideally, this allows the user to focus attention, as well as financial resources, on a few important factors in a process. They are created by plotting the cumulative frequencies of the relative frequency data (event count data), in descending order. When this is done, the most essential factors for the analysis are graphically displayed in an orderly format. An example of a Pareto Chart is shown in **Figure 5**.

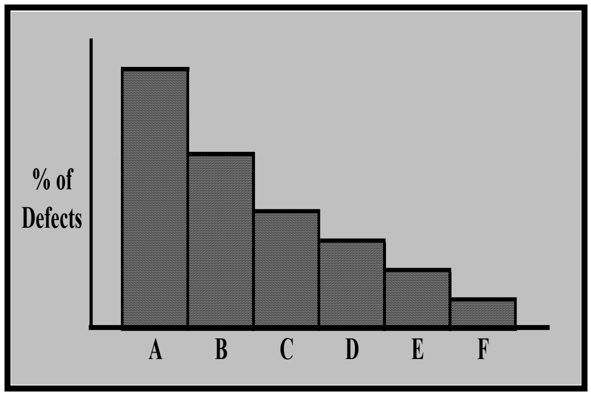


Figure 5: Pareto Chart

Another tool of quality that appeared to the team to have use in correcting quality problems at the Grayson plant was the ***Process Flow Chart.*** The Process Flow Chart is a pictorial representation describing a process under study. It may also be used to plan the stages of a project. Flow charts provide analysts with a common language or reference point when dealing with a project or process. Four particular types of flow charts have proven useful when performing process analysis: top-down flow chart, detailed flow chart, work flow diagrams, and deployment charts. Each of these different types of flow chart provides a different aspect of a process or task. Flow charts provide an excellent form of documentation for a process, and quite often are useful when examining how various steps in a process work together. Process flow charts can be used to analyze two separate stages of the process: The manufacture of the product and the finished product. When used to analyze the finished product or the operation of the process, flow charts use simple and easily recognizable symbols. Flow charting has been around for a very long time, and was used as long ago as 1921 (Gilbreth and Gilbreth, 1921). In fact, flow charts have been used for so long that no one individual is identified as the "Father of the flow chart". The reason for this is obvious. A flow chart can be customized to fit any need or purpose. Consequently, flow charts are recognized as an important quality improvement device. An example of a Process Flow Chart is shown below in **Figure 6**.

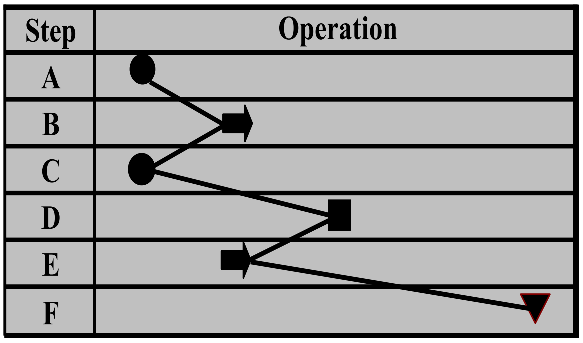


Figure 6: Process Flow Chart

The final tool of quality discovered by the team was that of ***Statistical Process Control Charts***. Control charts were originally devised by Dr. Walter Shewhart as a statistical approach to the study of manufacturing process variation for the purpose of improving the economic effectiveness of processes. This method is based on continuous monitoring of process variation. A typical control chart is a graphical plot of a quality characteristic, measured or computed from a sample, versus the sample number or time. The chart contains a center line that represents the average value of the quality characteristic corresponding to the in-control state. Two other horizontal lines, called the upper control limit (UCL) and the lower control limit (LCL) are also drawn. These control limits are usually set at $+3\sigma$ for the UCL and -3σ for the LCL. If the process is in control, 99.7% of sample points will fall between the UCL and the LCL. As long as the points plot within the control limits, the process is assumed to be in control, and no action is necessary. However, several points falling outside of the control limits, or certain patterns in the points, are evidence that the process may be moving out of control. If so, investigation and corrective action are then required to find and eliminate the special cause(s) of variation. The control points are connected with straight line segments for easy visualization. Even if all of the points plot inside the control limits, if they behave in a systematic or non-random manner, then this is an indication that the process is probably out of control. (Evans and Lindsay, 2005). An example process control chart is shown in **Figure 7**.

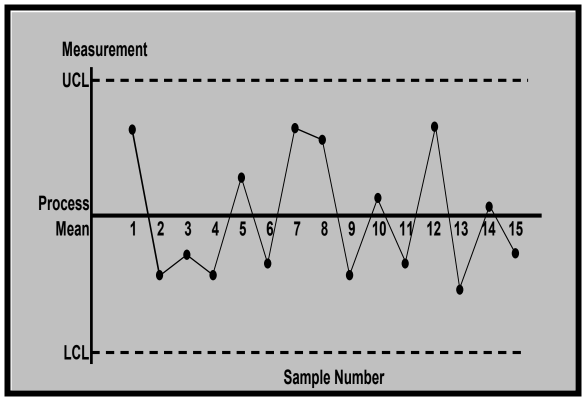


Figure 7: Example Process Control Chart

3. Methodology

In keeping with the importance of having a project champion, the team requested that Ms. Jacqueline Taylor, floor production manager at Grayson, serve as the interface between the team and upper management at the Grayson facility. Ms. Taylor was viewed as a particularly good choice in view of her extensive knowledge of production operations at Grayson. Ms. Taylor's duties were to attend team meetings, to provide any necessary guidance, and to keep upper management abreast of the team's progress.

As the team set out to acquire the critical information for employing the tools of quality needed to analyze production operations at the Grayson mill, it found that a substantial amount of data had already been collected from routine plant operations. In fact, this information was found to comprise five months of operation, from August through December of 2014, and in a form that made it readily usable with the tools of quality. This data was found to be expressed in terms of non-conforming product or NCP codes. Because sufficient data was already tabulated in the available files, it was not found necessary to take any measurements, but only to employ the sorting and pivot table features of EXCEL to develop the desired relationships among the data.

Based upon the production data discovered to be available, the team decided to revise its study plan, tailoring it to the Grayson application. This study plan was as follows:

1) Develop production process diagram; 2) Develop a cause and effect diagram to identify possible sources of NCP; 3) Using the available dataset, develop Pareto plots, and other plots as necessary, to determine which possible causes of NCP were actual, or root causes; 4) Develop and present a remedial plan to decrease NCPs to 10% or less; 5) Perform a limited implementation of preferred solution;

6) Monitor system to determine whether or not preferred solution(s) corrects problem(s); 7) Develop remedial plan to implement preferred solution; 8) Implement remedial plan for production operations; 9) Set up control chart for proportion defective, viz. a "p" chart, to monitor process(s) and ensure continuing compliance with correction(s).

The team then proceeded to apply this methodology to the problem of excessive NCP generation at Grayson.

4. Application of Methodology

Although the production machinery at Grayson was well maintained, it was somewhat dated. This caused the team to wonder whether excessive NCP might be due to a particular machine combination(s). To do this, the team began by drawing a process diagram for fabric manufacture at Grayson. This diagram is shown in **Figure 8**.

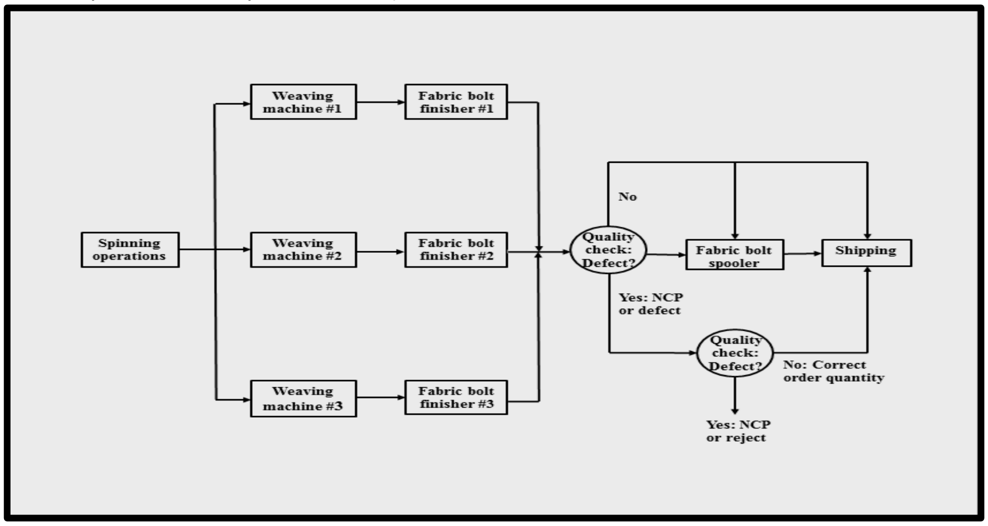


Figure 8: Process diagram for fabric production

It will be seen from **Figure 8** that the production process at Grayson consists of three stages: 1) Weaving; 2) Finishing (3); and bolt spooling. Examples of each type of machine are shown in **Figures 9, 10** and **11** respectively.

Figure 9: Typical weaving machine

Figure 10: Typical finishing machine

Figure 11: Typical bolt spooling machine

The process diagram enabled the team to examine NCP as a function of a particular machine combination(s), as shown below. Next the team developed an Ishikawa diagram to depict the possible causes of NCP generation. This diagram is shown in **Figure 12**.

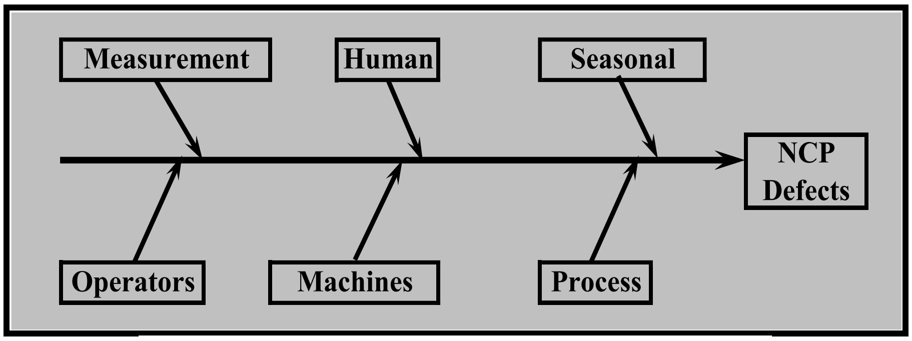


Figure 12: Ishikawa diagram for Grayson application

Having analyzed the manufacturing processes and possible sources of NCP generation, the team decided to examine NCP generation as a function of time. This was done by extracting the raw data for NCP generation by month, and constructing a preliminary Pareto diagram with the actual number of NCPs versus months. This diagram is shown below in **Figure 13**.

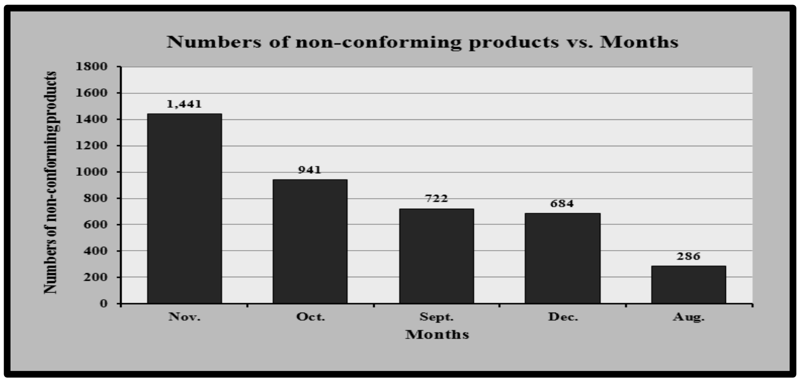


Figure 13: Preliminary Pareto diagram of NCP generation by month

The preliminary Pareto diagram was then converted to a formal Pareto diagram as shown in **Figure 14**.

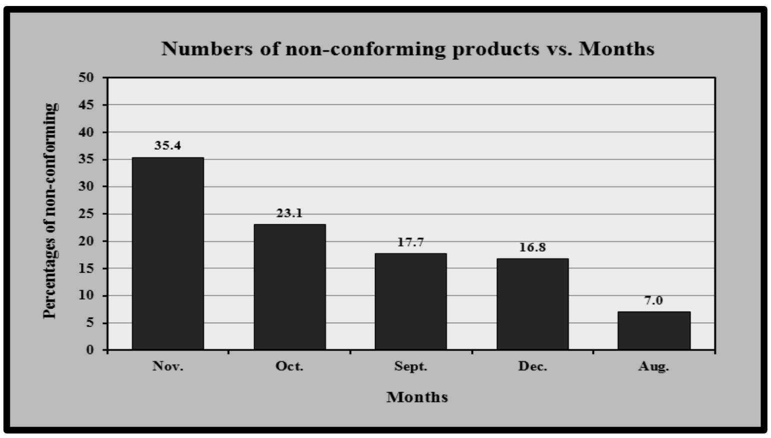


Figure 14: Formal Pareto diagram of NCP generation by month

It was also determined necessary to examine NCP generation as a function of the NCP category. This was done by extracting the raw data for NCP generation by NCP category, and constructing a preliminary Pareto diagram with the actual number of NCPs versus NCP category. This diagram is shown below in **Figure 15**.

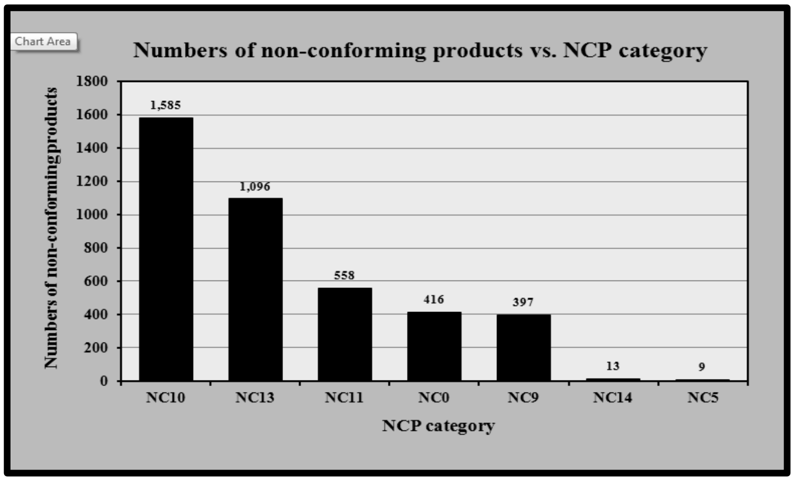


Figure 15: Preliminary Pareto diagram of NCP generation by category

The above preliminary Pareto diagram was then converted to a formal Pareto diagram as shown in **Figure 16**.

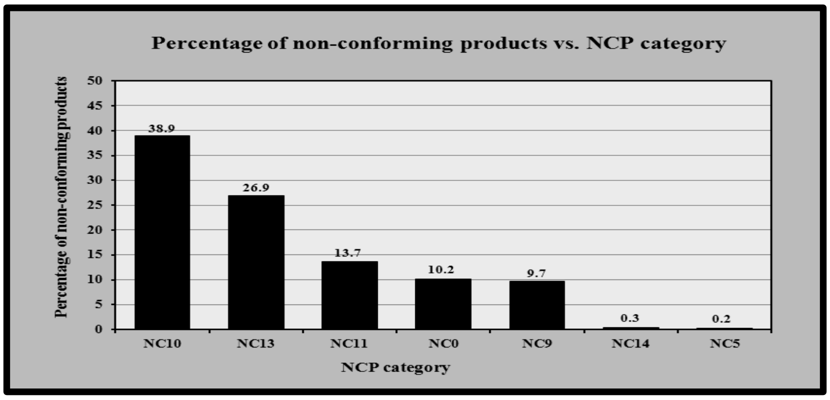


Figure 16: Formal Pareto diagram of NCP generation by category

Next, it was necessary to examine the extent to which operator error was responsible for NCP generation. Thus, the team next extracted the raw data for NCP generation and completed another preliminary plot of NCP generation as a function of operator error. This preliminary plot is shown below in **Figure 17**.

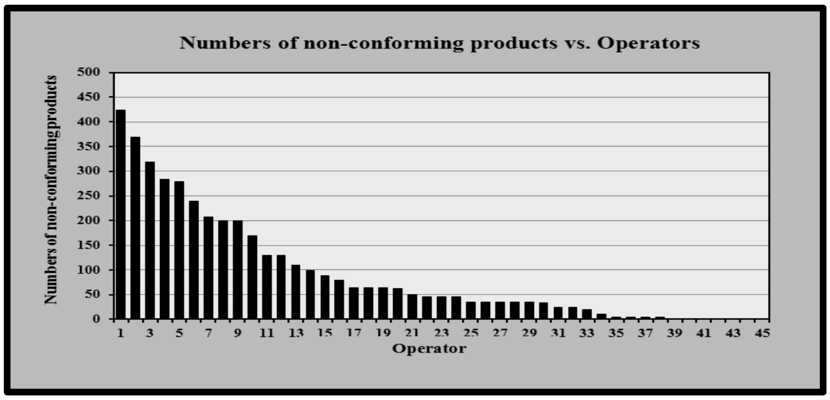


Figure 17: Preliminary Pareto diagram of NCP generation by operator

As before the preliminary, raw data plot was converted to a formal Pareto diagram as shown in **Figure 18**.

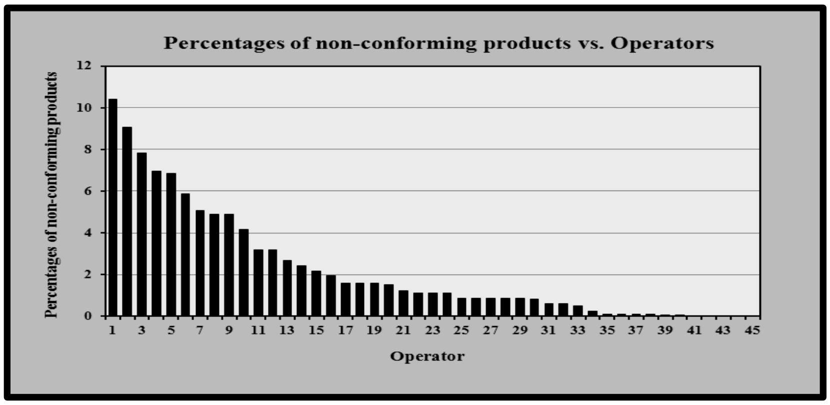


Figure 18: Formal Pareto diagram of NCP generation by operator

Next, the team extracted the data for NCP generation as a function of the production machinery, and prepared the plot shown in **Figure 19**.

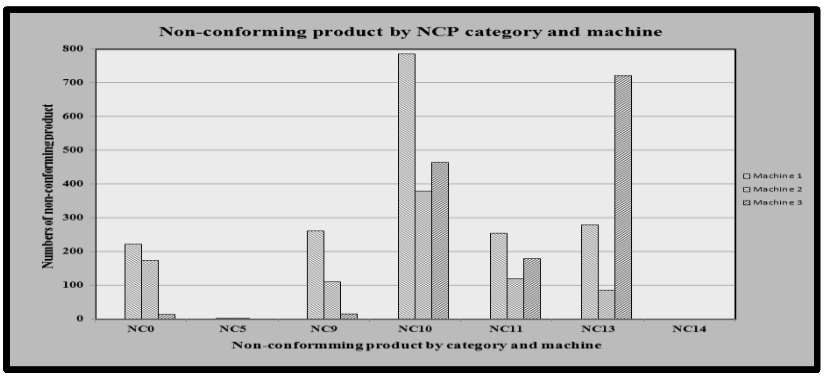


Figure 19: NCP generation by machine

The team realized, of course, that any results as to NCP generation by machinery were attributable to each combination of weaving and finishing machines.

The next question to be addressed, by the team was whether the data indicated any correlation between NCP generation and a particular shift(s). The next stage of the analysis, therefore, was to extract data portraying the relationship between NCP generation and shift. The plot of this data is shown in **Figure 20**.

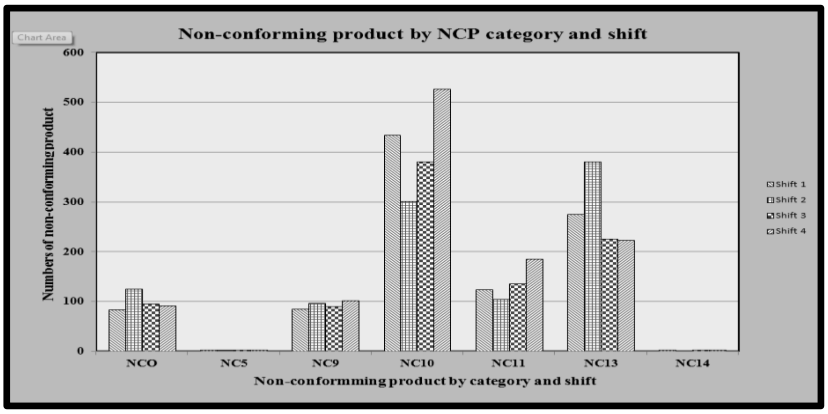


Figure 20: NCP generation by category and shift

The next relationship to be examined by the team was that of NCP generation as a function of NCP type and month. The plot of this data is shown in **Figure 21**.

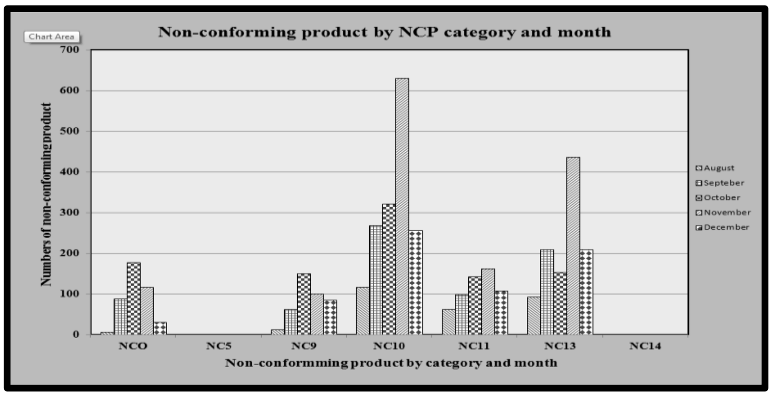


Figure 21: Numbers of NCP versus NCP category and month

Last, the team prepared a control chart in anticipation of the necessity for long-term monitoring of the process. Since Grayson specified the maximum NCP as a percentage of the maximum monthly inventory, it was decided to use a proportion defective, or "p chart" (Russell and Taylor, 2011). The maximum desirable NCP specified by Grayson was 10%. Thus, the upper control limit (UCL) was set at 10%, corresponding with three standard deviations above the mean. The lower control limit (LCL) was then set at 0%. The mean, p-bar, was set at 5%. The control chart plot for NCP versus months is shown in **Figure 22**.

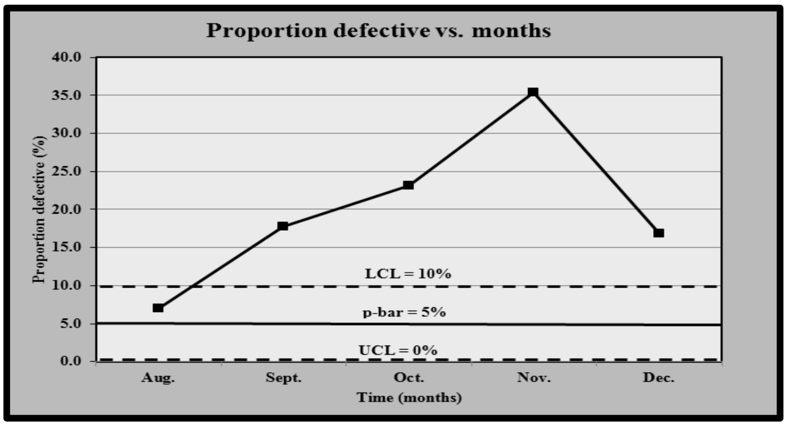


Figure 22: Control chart for NCP versus months

The foregoing analysis disclosed some very important issues in production operations at Grayson. These issues are discussed in the following section.

5. Results

66% of all NCP generated was of types NC10 and NC13. The team found it curious that the total of these two NCP categories was nearly the same as the total for the months of November and October, as previously discussed for **Figure 14**.

The Pareto diagram for NCP versus operator, **Figure 18**, was equally remarkable, for it shows that operators 1 through 9, viz. 20% of the operators, generate almost 62% of the total NCP. Further, operators 1 and 2 generate almost 20% of the NCP all by themselves. At the other end of the spectrum, operators numbered 35 through 45 generated a cumulative NCP of less than 0.5%, with operator 45 having generated no NCP at all during the five month period.

Figure 19, NCP type by machine, was very illuminating. First of all, no NCP of types NC1, NC2, NC4, NC6, nor of NC12 was generated at all. Also, types NC5 and NC14 were nearly zero. Maximum NCP was of type NC10

The Pareto diagram for NCP versus month of generation, **Figure14**, disclosed a remarkable situation: Nearly 60% of the total NCP for the first five months was generated in the months of November and October with November being nearly 35% larger than October. Moreover, more than a third of the total was generated in the month of November alone.

Figure 16, shows the relationship between percentage of NCP and NCP category. This plot shows that nearly

generated on machine #1, an amount of 786, with machine #3 following closely at 722 of type NC13. Machine combination #1 generated the most NCPs, an amount of 1,806, with machine combination #3 generating the second greatest at a total of 1,394. Machine combination #2 generated the least number of NCPs, a total of 874.

The generation of NCP by shift is shown in **Figure 20**. As was the case with NCP generation by machine combination, the largest category by shift was once again NC10. Also, NC 13 was again the second largest class. Generation of categories NC5 and NC14 was also once again small and practically non-existent.

Figure 21 shows NCP generation by month and NCP category. The same comments made for NCP generation by shift are also true of NCP generation by month. It may be said in addition for category NC10 that NCP generation for each month was greater than that for the comparable month in any of the other categories. This was also true of the second largest NCP category, NC13, except for the month of October.

As might be expected the control chart for proportion defective, the p-chart in **Figure 22**, shows NCP generation at the Grayson to be entirely out of control. The recommendations made by the team in the following section were aimed at bringing continuous process improvement to bear in correcting the excessive NCP generation problem at Grayson.

6. Recommendations

From the results obtained by the team, it was evident that immediate and robust measures were required to correct the problem of excessive NCP generation at Grayson.

The fact that NCP generation was so much greater in November, and to some extent October, suggests that some sort of special cause was present in November, and to a lesser extent October. For the second order of investigation, therefore, the team recommended that an examination be conducted of the production records for October and November to determine what special cause(s) might have entered the production process to account for the excessive NCP generation.

The team was gratified to find that with regard to NCP generation, they had approximately verified the Pareto principle, viz. that 20% of the operators were responsible for approximately 60% of the NCP generation. To correct this problem, the team recommended as a short term measure that these employees receive additional supervisory scrutiny to determine the cause(s) of their excessive NCP generations. It was also recommended that those employees generating the least NCPs receive additional oversight so as to learn why their work was of higher quality. As a longer term measure, the team recommended a training program to improve operator skills.

The team found it curious that of the12 possible types of NCP, 8 were not generated at all, and 2 more were nearly 0. By far the largest generated was NC10 from machine combination #1, with NC13 a close second from machine combination #3. It was further curious that the number of NC10 generated on machine combination #1 was nearly three times that of type NC13. In addition, the number of NC13 generated on the machine #3 combination approached twice that of the NC10 that was generated. This peculiarity suggested that special causes of variation were present in the production process. The team recommended, therefore, that further analysis be conducted to identify those causes.

Examination of the NCP generation by shift disclosed that shifts 1, 3, and 4 generated more NC10 than any other NC category. Shift #2 generated more NC13. The team observed that NC10 and NC13 were also the greatest NCPs when examined by machine. Since all operators were rotated over all shifts, the team concluded that a substantial part of the NCP generation must likely be due to machines #1 and #3, or to the operation of those machines. This also appeared possible to a lesser extent on machine #2. This, along with the fact that 9 of the operators were collectively generating approximately 60% of the total NCP, led the team to recommend that the second order analysis be aimed at determining which operators were generating each category of NCP.

Examination of NCP categories by month disclosed what the team had already suspected, viz, that the greatest amounts of NCP were of types NC10 and NC13, and were generated in November. However, the totals of NC10 and NC13 for September, October and December were nearly equal and quite large. This strengthened the team's earlier conclusion about special causes of variation having entered the production process, but also indicated that they might have begun earlier than previously thought. This discovery gave further impetus for the first recommendation.

Finally, in keeping with decreasing NCP generation to an acceptable level, the team had begun to monitor the NCP by constructing p-charts to monitor proportion defective. The team recommended expanding this initial step to a complete quality control program. Along with this, it was recommended that Grayson adopt a continuous process improvement program (CPI). An economic analysis estimated that a CPI program would save Grayson as much as $45,000 per month in working capital and $9,000 per month in inventory costs. As a comprehensive, corporate measure, it was recommended that as soon as these programs were well established and functioning properly at Grayson, that they also be implemented in the remainder of the Wilmington Textile's subsidiaries. These recommendations are summarized in **Table 1** below.

NCP Issue	Recommendation
1. Greatest NCP generation in October and November	**Examine production records for October and November to investigate special causes of variation**
2. 60% of NCP generation by 20% of operators; Pareto principle	**Temporary increased supervision of employees generating the greatest NCP**
3. Largest generation of NC 10 on machine combo 1, and NC 13 on machine combo 2	**Study and isolate special causes of variation on machine combinations 1 and 2**
4. NC 10 and 13 greatest over all machines and all shifts	**Perform second order analysis to determine which operators are generating the NCPs**
5. Greatest NC10 and 13 in month of November	**Same as recommendation for #1 to examine special causes specifically for NC10 and 13**
6. Decreasing NCP generation to an acceptable level	**Implement CPI program with p-charts to monitor processes at Grayson and then in subsidiaries**

Table 1: Summary of NCP issue and recommendation

7. References

Ahire, J. L. (1997). *The Management Science-Total Quality Management Interfaces. Interfaces.* The Institute for Operations Research and the Management Sciences (INFORMS). Hanover, MD

Bousquet, G.-H. et G. Busino (1964). *Vilfredo Pareto: Cours d'Économie Politique*. Nouvelle édition par Librairie Droz. Geneva, pages 299–345

Deming, W.E. [1986]. *Out of the Crisis*. MIT Press. Cambridge, MA, page 88

Evans, James R., and William M. Lindsay (2005). *An Introduction to Six-Sigma*, Thompson Southwestern, United States, et al, pgs. 375

Gilbreth, Frank B. and Lillian M. Gilbreth (1921). *Process Charts*. American Society of Mechanical Engineers.

Ishikawa, Kaoru (1968). *Guide to Quality Control.* Tokyo: JUSE.

Ishikawa, Kaoru (1976). *Guide to Quality Control.* Asian Productivity Organization.

Ishikawa, Kaoru (1986), *Guide to Quality Control*. Tokyo: Asian Productivity Organization. 2 ed. p. 41.

Pearson, Karl (1895). *Contributions to the Mathematical Theory of Evolution*. *II. Skew Variation in Homogeneous Material*. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, p. 186.

Russell, Roberta S. and Bernard W. Taylor (2011). *Operations Management*, John Wiley & Sons, Inc., 5th edition, Upper Saddle River, NJ

Schultz, John R. (2006). Measuring Service Industry Performance: Some Basic Concepts. *International Society for Performance Improvement*. p. 3.

Shewhart, W. A. [1939]. *Statistical Method from the Viewpoint of Quality Control*. Department of Agriculture. Dover, 1986, page 45

Stephens, K. S. (1965a). Quality Control – Small and Large. Proceedings of the Quality Conference, Department of Commerce, Bureau of Public Roads, Washington, D.C.

Stephens, K. S. (1965b). "Process Capabilities – Small and large." Transactions of the 1st Pan American Congress on Quality Control and Statistics in Industry. Mexico City, Mexico

Sutterfield, J. S. and Conrod S. J. Kelly (2005). Metrics for Continuous Process Improvement. *2005 IEMS Proceedings.* Cocoa Beach, FL. pgs. 370-378

Tague, Nancy R. (2004). The Seven Basic Quality Tools. *The Quality Toolbox*. Milwaukee, Wisconsin. American Society for Quality. p. 15.

A Veteran's View: Comfortability Using Avatars to Express Emotion

Joan M Savage
Florida Institute of Technology, College of Aeronautics
savagej2014@my.fit.edu

Abstract

Virtual reality (VR), augmented reality (AR) and video games (VG) are growing in popularity and in today's society are fulfilling and satisfying genuine human needs that the real world is currently unable to satisfy. They are inspiring, teaching, and engaging us in ways that reality is not [1]. High veteran suicides are attributed to Post-Traumatic Stress Disorder (PTSD) and isolation. This research explores developing ways to meet genuine human needs through the use of avatars in virtual environments (VR and AR) to provide confidentiality, privacy, safety and most importantly community to express real-time emotions without fear of any repercussions or personal identification.

1. Introduction

Twenty-two veterans commit suicide a day [2]. This number does not include *attempted* or *failed* suicides. Eleven percent of those who attempted suicide in 2009 (and did not die as a result of this attempt) made a repeat suicide attempt [3]. Some veteran's advocacy groups have filed a class-action lawsuit claiming that the Veteran Health Administration (VHA) is not providing adequate and timely access to mental health care, and this has led to an "epidemic of suicides" [4]. A press release has been issued stating that mental health of veterans was now the highest priority for the Department of Veterans Affairs (VA) [5] and recent research on veterans exposed to combat indicates that they are at increased risk for mental health disorders (especially PTSD) and depression) which are major factors in suicide [6]. Although there are successful treatments for full-blown PTSD, early interventions are lacking. An early intervention method; a visuospatial computer game (e.g. "Tetris") has shown to interfere with flashbacks [7].

The Center for Disease Control (CDC) is working with the VA to better understand the magnitude of the problem of suicide among veterans [8]. Responsibility for prevention of veteran suicide lies primarily within the VA. Such a challenging task necessitates collaboration with other federal agencies, state and local governments and other organizations like Florida Institute of Technology, and others like private funding [9]. Existing strategies for connecting military personnel to these services are inadequate. For example, Operation Iraqi Freedom veterans were associated with high utilization of mental health services after deployment. This high rate highlights challenges in ensuring that there are adequate resources to meet the mental health needs when they return. But most of these veterans were not even captured by existing screening procedures[1] and *more than* 50% of those referred for a mental health reason were documented as needing to receive follow-up care but *less than* 10% were referred through the screening program [10].

1.1. Problem Statement

There are three main problems that affect veterans concerning receiving treatment using VA health care services - Eligibility: Not all veterans are eligible for VA health care services. Enrollment: Not all eligible veterans actually enroll in VA health care services. Use: Not all enrolled veterans want to use VA health care services [11]. Currently, there are nearly 22 million veterans living in the United States, *only* 5.9 million actually use VA health care services; therefore, 16 million veterans are NOT receiving health care from the VA [12] but depending on charity and local and state resources.[2]

Regardless of how injured (mentally or physically), the current wait time is over 125 days *just to have their medical records reviewed* [13] for service-connectedness (an injury sustained in the military).

Once the veteran has demonstrated service-connectedness and their disability compensation claim is approved, the next step is registering with a local VA Medical Center to request health care services. Currently, there are long wait times to be seen by a primary care physician; then long wait times if the veteran is referred to a specialist; for example, to psychiatry.

Mechanisms to support veterans while they are in transition and waiting to receive professional treatment are needed. And if the veteran refuses to seek psychiatric help from the VA (to avoid a PTSD label) due to professional reasons.

Including the many reasons listed above, there is a substantial burden of mental illness experienced by many individuals in the military and an alarm has been sounded for innovative ways to deliver services to reduce this burden; [14] the Surgeon General's 2012 National Strategy for Suicide Prevention (Objective 2:3) supports innovative applications like VR and VG [15].

2. Methodology

To better understand the avoidance of seeking mental health help, interviews and observations were conducted on veterans (with law enforcement connections). This study had a small N of five Subjects meeting the standards for Nielsen's recommendations [16].

2.1. Critical Decision Method Interview (CDM)

Five unofficial interviews were conducted with United States military veterans. For CDM to be most effective for this type of research questioning, it is imperative that participants felt *safe* to share explicit details. Every possible assurance was given that their identity and shared information would remain confidential and anonymous. Another concern for participants with a security clearance; official documents like a written informed consent would require approval from respective departments and authority heads before signing. In the interest of maintaining an atmosphere of trust and safety with participants, nothing identifying will be published. These interviews were conducted solely for the purpose of developing a needs analysis.

3. Results

An advantage of using CDM, the open-ended questions allow veterans to experience the freedom to express themselves without social tension, fear of repercussion or judgement and allowed the space to tell their stories.

[1] Screening procedures occur before a soldier is officially discharged from military service. Service members must affirm health to be released from military service.

[2] According to www.va.gov, it should be noted that Florida ranks number three in veteran population.

Demographics included four males and one female ages ranging from mid-20 to early 50's for different perspectives over different wartimes but all served in the United States Military for an extended period of time.

Subjects One and Two are *currently* working in law enforcement post-military. Subject Three is not only currently working in law enforcement but also worked as a Military Police Officer while serving within the military. Subject Four dealt with civil affairs and law enforcement while serving in Afghanistan and plans to work within the law enforcement field (i.e. Federal Bureau of Investigation, Sheriff's Department, etc.). Subject Five was discharged from the military under conditions not honorable for reasons (according to the veteran) that were unjust. Subject Five clearly affirmed feelings of wrongful termination by the military justice system. However, due to this type of discharge the veteran is automatically ineligible to receive medical treatment from VA Medical Centers and in most federal situations lost the status of US military veteran. However, Subject Five disclosed the need for psychiatric treatment.

Four of the Subjects expressed fear of losing the ability to carry a concealed weapon, and the fear of losing a security clearance, or the inability to maintain or retain one with an official PTSD diagnosis.

All participants claimed to personally suffer from PTSD, as well as personally know other currently employed law enforcement officers who suffered from undiagnosed PTSD.

Subject Three has been in authority over other law enforcement personnel (military and civilian) and emphatically stated that through personally observing and through unofficial disclosure these officers were dealing with PTSD ranging in all levels of severity.

Two of the Subjects discussed suicide ideation. But all Subjects remained untreated due to the potential repercussions and risks to their professional field. And all Subjects remained untreated for fear of being stereotyped as *crazy*. Currently, law enforcement has an elevated risk of suicides [18].

3.1. Operation Safe-Space

This needs analysis supports the need for using an avatar within a virtual online space where individuals can meet other veterans for peer-to-peer mentoring, counseling, and socializing by chatting in real-time with real-time emotions; an avatar will be used to protect the identity and privacy by offering anonymity, safety and confidentiality so **anyone** can seek help at **any time** without repercussions.

The avatar mimics user's real-time emotion in honest emotion capture yet providing privacy. Privacy is vital for individuals to feel safe enough to share real emotions and feel free to express all concerns including PTSD and suicide ideation. This private, confidential and safe online environment using an avatar that looks nothing like the real user is called Operation Safe-Space (OS2).

Using a personal avatar designed to express the emotional state of the veteran using current available software (Figure 1) and 3D cameras, Kinect, etc. (displaying the veteran as an avatar) which *reads* the emotional expression of the veteran in real-time. Operation Safe-Space not only provides a safe environment to share genuine emotion but also provides a secondary factor of teaching a veteran to express emotion *accurately* and *thoughtfully* to help the transition to civilian life.

Figure 1. Brekel (taken from Google Images)

"Blank-Face Syndrome" (BFS) [19] is a condition a veteran may deal with when lacking proper and *thoughtful* facial expressions in human-to-human interactions. This is damaging to relationships and damaging within social climates but necessary in military operations and law enforcement.

The user can cue the avatar to their real-time true emotion by using self-observation. The veteran would not only be helping the avatar assess their emotions correctly but also help their own understanding and assessment of their true feelings; self-emotional identification.

The avatar will help the veteran be "thoughtfully expressive [20]" when a veteran is clued that their expression is not matching the emotion they believe they are feeling *because* the avatar (software) is not accurately reading/displaying the intended emotion. This will help veterans acclimate to civilian life and civilian relationships by teaching proper emotional expressions. Acclimation to civilian living also fundamentally requires a social circle.

Real world ethnography or cognitive ethnography employs traditional ethnographic methods to build knowledge of a community of practice and then apply this knowledge to the micro-level analysis of specific episodes of activity. The principal aim of cognitive ethnography is to reveal how cognitive activities are accomplished in real-world settings. Cognitive ethnography is a particularly apt method for studying instruction in both formal and informal settings [21].

Emerging evidence suggests that the potential of social networking applications to improve health outcomes lies not only in their ability to provide health information but also in their social and supportive functions that may appear absent in the traditional health care system [22]. Social interaction is invaluable for those struggling with suicide ideation and social support seems to have a protective effect on Orbitofrontal (OFG) volume in the brain. For combat veterans, greater combat experiences are associated with reduced OFG volume and the OFG regions have been proposed to play a role in object identification, learning of stimulus-reward associations and value-based decision-making. Volume and functioning may therefore be important for recognizing the significance of stimuli (particularly rewarding stimuli) to produce appropriate responses [23]. As a veteran transitions into civilian life, appropriate emotional responses are detrimental for their successful acclimation; building their social circles, and the development of healthy functional relationships. Social circles and healthy functional relationships are vital for suicide reduction and overall improved mental health.

Although combat experience is a major contributor to PTSD and suicide ideation - another example are veterans who are victims of military sexual trauma (MST) and the long-term emotional consequences associated with MST.

The problem with reporting a sexual assault in the military is that the veteran is diagnosed with PTSD.

Reported sexual assaults in the military increased by over 50 percent [24]. Operation Safe-Space can offer individual therapy, peer-to-peer counseling, community groups and social support at any time of the day for extended periods of time.

Veterans can communicate the atrocities of sexual abuse from the safety of their own home and use an avatar to protect their identity. Computers do not make judgments or look down on people the way another human might; therefore, an avatar is an easier way to share private information [25]. Because the veteran will be communicating and connecting with *real* humans they are accessing and building healthy social environments and relationships using a computer-generated avatar.

3.2. Affordances and Enactment

Using two guiding principles of affordances and enactment as discussed in the Usability Engineering course at Human-Centered

Design Institute: affordances (suggestive actions) which includes values (either positive or negative) attached to actions toward an entity (avatar) that are to be visually discovered (perceived) within their environment depending on their needs and within their shared community – enabling socialization & social interactions. And enactment (act something out): in open complex virtual environments using an avatar and experienced through action depending on the environment and the cultural context but still considering the usability output of product adoption and diffusion. Ultimately, the veteran must feel comfortable and safe using the avatar to connect with other people virtually. The veteran must be comfortable expressing emotion using the avatar and be willing to connect socially and emotionally with a community.

The word emotion comes from Latin *motere* that means movement. Thus, emotion is immediately linked to action, and each basic or fundamental emotion has a vital role (function) [26]; therefore, in order for the veteran to learn to behave appropriately, the veteran has to perceive information that enables understanding of the current situation, but also to anticipate the evolution of the situation [27]. Operation Safe-Space can offer situations designed specifically for the veteran's current situation, cultural background and historical evolution.

4. Conclusion

Soldiers are released from the military without real-world ability to cope with being a civilian and connecting with other civilians. Many veterans understand suicide in the abstract but when asked how to identify and handle a person contemplating suicide they could not [28]. Current socio-technical systems designed to teach soldiers/veterans about suicide prevention is failing. Figure 2 is the initial specifications for training methods and mapping for suicide identification [29] for caregivers at VA Medical Centers identify individuals contemplating suicide.

There is still a reactive approach (react to the attempt of suicide) verses a proactive approach. The progression to suicide is non-linear.

All Subjects interviewed admitted they exhibited PTSD thoughts, actions and behaviors (even suicide-ideation) but were unwilling to pursue professional help due to the possible repercussions. All Subjects showed positive responses concerning using Operation Safe-Space for treatment and help since it offered privacy, confidentiality and safety. All Subjects agreed an avatar was the best choice for privacy; offering real-time expression of emotion safely and anonymously.

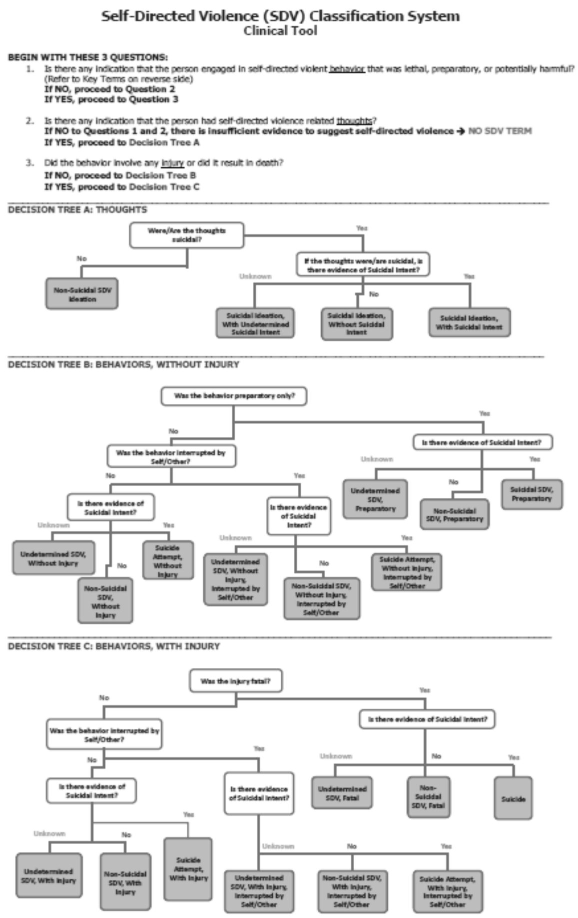


Figure 2. Classification system

5. Future Applications

All five senses can be activated to support emotion regulation which using an avatar, this methodology could influence veterans preferred behavior for managing stress [30], and help change behavioral patterns of avoiding necessary treatment for fear of repercussions. Smell is especially powerful for positive and negative associations.

Emotion regulation supports the extrinsic and intrinsic processes responsible for monitoring, evaluating, and modifying emotional reactions; with extrinsic processes being external to the veteran, i.e. vanilla smell can be soothing as a peer within the VR environment helps soothe anxiety; or intrinsic processes where the veteran internally regulates as the smell of vanilla is used simultaneously for positive association. Emotion regulation is, therefore, multi-dimensional and complex, and refers to more than just modifying, or controlling, emotional reactions [31]. A complete supportive system could be in place to help the veteran in day to day living as a civilian and learning to handle normal day-to-day emotion regulation and social interaction.

Currently, there are efforts underway for long-term space travel; including colonizing Mars. Operation Safe-Space methods are invaluable for the crew as they perform their day-to-day repetitive reports and activities. Mental health, emotion regulation, social interaction, and sensory stimulation is vital to the mission's long-term success. The issues identified with veterans can translate to the crew during this long-term space travel and during colonization. It can be used in any isolated or closed environment (even submarines) especially where all their activities and actions are constantly monitored and controlled. Operation Safe-Space is essential for the mental health and well-being of the crew.

Soldiers are consistently exposed to a lack of privacy, and an invasion of space. As a result, once released from the military a veteran may pursue individual privacy and space; which can lead to isolation. Isolation is a major contributor in suicide ideation [32]; therefore, Operation Safe-Space offers a safe and solid recommendation to the potential spiraling degradation of mental health especially over long periods of time.

6. References

[1]. McGonigal J. (2011). Reality Is Broken: Why Games Make Us Better and How They Can Change the World. London: Penguin.

[2]. Kemp, J. & Bossarte, R. (2012). Retrieved from: http://www.va.gov/opa/docs/suicide-data-report-2012-final.pdf

[3]. Bossarte, R. & Kemp J. (2011). VA Program for Suicide Prevention; Retrieved from: http://www.ihs.gov/suicidepreventionsummit/alaska/documents01/VAprogramBossarteKempPresentation.pdf

[4]. Sundararaman, R., Panangala, S. & Lister, S. (2008). CRS Report for Congress; Suicide Prevention Among Veterans. Domestic Social Policy Division (US); Order Code RL34471. Retrieved from: http://www.fas.org/sgp/crs/misc/RL34471.pdf

[5]. Shinseki, E. (2012). President Obama Signs Executive Order to Improve Access to Mental Health Services for Veterans, Service Members, and Military Families [News Release]; (August 31, 2012).

[6]. Sundararaman, R., Panangala, S. & Lister, S. (2008). CRS Report for Congress; Suicide Prevention Among Veterans. Domestic Social Policy Division (US); Order Code RL34471. Retrieved from: http://www.fas.org/sgp/crs/misc/RL34471.pdf

[7]. Holmes, E., James, E., Coode-Bate, T. & Deeprose, C. (2009). Can playing the computer game "Tetris" reduce the build-up of flashbacks for trauma? *Cognitive science*. Retrieved from: http://www.plosone.org/article/info:doi/10.1371/journal.pone.0004153

[8]. Centers for Disease Control and Prevention, (2011). Preventing Suicide Program Activities Guide. Retrieved from: http://www.cdc.gov/violenceprevention/pdf/PreventingSuicide-a.pdf

[9]. Bagalman, E. (2012). Suicide Prevention Efforts of the Veterans Health Administration Congressional Research Service (US); (2012, Feb.), Report No.:R42340.

[10]. Hoge, C. W., Auchterlonie, J. I., & Milliken, C. S. (2006). Mental health problems, use of mental health services, and attrition from military service

after returning from deployment to Iraq or Afghanistan. *Journal of the American Medical Association, 295*, 1023–1032. doi:10.1001/jama.295.9.1.1023

[11]. Bagalman, E. (2012). Suicide Prevention Efforts of the Veterans Health Administration Congressional Research Service (US); 2012 Feb. Report No.:R42340.

[12]. Bagalman, E. (2014). The Number of Veterans That Use VA Health Care Services: A Fact Sheet. [June 3, 2014]. Retrieved from: http://fas.org/sgp/crs/misc/R43579.pdf

[13]. Heller, D., Moran, J., Vitter, D., Casey, JR., Heinrich, M. & Tester, J. (2014). Retrieved from: http://www.casey.senate.gov/download/va-backlog-march-2014-report

[14]. Kazdin, A. E., & Blasé, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on Psychological Science, 6*, 21-36, doi: 10.1177/1745691610393527.

[15]. U.S. Department of Health and Human Services. (2012). Office of the Surgeon General and National Action Alliance for Suicide Prevention. National Strategy for Suicide Prevention: Goals and Objectives for Action; 2012.

[16]. Nielsen J. (1994). Estimating the number of subjects needed for a thinking aloud test. International Human Computer Interaction. 1994; 41, 385-397.

[17]. USA.gov (nda). Retrieved from: https://www.usa.gov/laws-and-regulations

[18]. Violanti, J.M., Robinson, C.F. & Shen, R. (2013). Law enforcement suicide: a national analysis. Retrieved from: http://www.cdc.gov/niosh/nioshtic-2/20044781.html

[19]. Savage, J. (2014). A Veteran's Guide to Civilian Living. Retrieved from: http://www.amazon.com/A-Veterans-Guide-Civilian-Living/dp/1502753685

[20]. Savage, J. (2014). A Veteran's Guide to Civilian Living. Retrieved from: http://www.amazon.com/A-Veterans-Guide-Civilian-Living/dp/1502753685

[21]. Williams, R. F. (2006). Using cognitive ethnography to study instruction. *Proceedings of the 7th International Conference of the Learning Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

[22]. Gibbons, M.C., Fleisher, L., Slamon, R. E., Bass, S., Kandadai, V. & Beck, J. R. (2011). Exploring the potential of web 2.0 to address health disparities, *Journal of Health Communication, 16* (1), 77-89, http://dx.doi.org/10.1080/10810730.2011.596916

[23]. Aupperle, R.L., Connolly, C.G., Stillman, A.N., May, A.C. & Paulus, M.P. (2013). Deployment and Post-Deployment Experiences in OEF/OIF Veterans: Relationship to Gray Matter Volume. Retrieved from: http://www.cdc.gov/niosh/nioshtic-2/20044781.html

[24]. Cooper, H. (2014). Pentagon Study Finds 50% Increase in Reports of Military Sexual Assaults. *The New York Times*, (2014, May 01). Retrieved from http://www.nytimes.com

[25]. Lucas, G., Gratch, J., King, A. & Morency, L. (2014). *Computers in Human Behavior.* It's only a computer: Virtual humans increase willingness to disclose; August, 2014 (37), 94–100. doi:10.1016/j.chb.2014.04.043

[26]. Stephane, L. (2007). Cognitive and emotional human models within a multi-agent framework. European Institute of Cognitive Sciences and Engineering. Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.8105&rep=rep1&type=pdf

[27]. Stephane, L. (2007). Cognitive and emotional human models within a multi-agent framework. European Institute of Cognitive Sciences and Engineering. Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.8105&rep=rep1&type=pdf

[28]. Savage, J. M. (2015). Usability Assessment of a Suicide Intervention-Prevention Mini-Game. *International Conference on Human-Computer Interaction,* (August, 703-708). Springer International Publishing.

[29]. Department of Veteran's Affairs. Retrieved from: http://www.mirecc.va.gov/visn19/docs/Clinical_tool.pdf

[30]. Han, S., & Pistole, M. C. (2014). College student binge eating: Insecure attachment and emotion regulation. Journal of College Student Development, 55(1), 16-29. Retrieved from: http://search.proquest.com/docview/1500821910?accountid=27313

[31]. Han, S., & Pistole, M. C. (2014). College student binge eating: Insecure attachment and emotion regulation. Journal of College Student Development, 55(1), 16-29. Retrieved from: http://search.proquest.com/docview/1500821910?accountid=27313

[32]. Savage, J. M. (2015). Usability Assessment of a Suicide Intervention-Prevention Mini-Game. *International Conference on Human-Computer Interaction,* (August, 703-708). Springer International Publishing.

.