ISSN 1750-9823 (print)
International Journal of Sports Science and Engineering
Vol. 02 (2008) No. 03, pp. 131-143

Evaluating the effects of aging for professional football players in combine events using performance-aging curves

William A. Young II [+] and Gary R. Weckman

Department of Industrial and Systems Engineering, Russ College of Engineering and Technology, Ohio University, 270 Stocker Center, Athens, Ohio 45701-2979, USA

(*Received May 28 2008, accepted July 3, 2008*)

Abstract. This paper presents a new methodology to overcome issues with proprietary, limited, or censored data. The methodology uses an artificial neural network (ANN) to translate a player's rating values to NFL Combine values. Though the Combine values typically are recorded once during a professional's career, the rating values exist well past a player's rookie year in professional football. After the ANN is trained, additional Combine performance values are forecasted using the rating database. Once the forecasts are made, performance-aging curves are applied to the aggregated data to investigate the impact that aging has on a player's expected Combine performance. The results of this methodology are somewhat consistent with other related literature in term of determining peak-performance age and rates of improvement; however, as expected, the rates of decline after peak are higher in football than other sports that are far less physically natured.

Keywords: performance-aging curve, rating system translation, artificial neural networks

1. Introduction

In 1982, the National Football League (NFL) held its first National Invitational Camp, which is more commonly referred to today as the NFL Scouting Combine. This event invites top draft-eligible players from the collegiate ranks to participate in a multi-day affair, which allows NFL personnel the opportunity to ask players personal and football related questions, review their medical history, and evaluate their physical skills in non-contact football drills. The participation in the Combine has doubled since the inaugural year, which now includes over 300 athletes annually (National Invitational Camp, 2007). The Combine events help executives make decisions about whom they would like to select in the NFL Draft, which allows professional teams to acquire the rights to players from colleges for their professional teams.

The decisions made by organizations during the Draft are not trivial. Teams spend a significant amount of resources to acquire the contractual rights of the players that are eligible. Scouts and other hiring professionals attempt to evaluate individual skills and identify how those skills will fit with their existing team. Correctly hiring players for the team has considerable ramifications to the team's ability to obtain large revenues.

Several characteristics are evaluated when making a hiring decision for the NFL. The largest proportion of this evaluation is placed on a player's performance during previously played football matches in their careers. However, the dispersion of talent varies greatly within the collegiate ranks. The level of talent can fluctuate on a particular team, with a team's opponents, or even a team's non-opponents. All of these factors make the evaluation process extremely difficult. Football is considered by some as the "ultimate team sport." Problems can arise when evaluating players during a game because players' performances are influenced by the playing abilities of their and opposing teammates.

A smaller, but still significant, portion of the evaluation process is placed on a player's individual non-contact physical performance. The NFL Combine provides the medium for this particular investigation. One benefit is that all subjects are tested on the same field with the same measuring equipment and standards. This practice reduces the possibility of data collection errors and provides a forum where all NFL organizations are given equally opportunity to evaluate available talent. The non-contact drills that players

[+] Corresponding author. Tel.: +1 740-593-1548; fax: +1 740-593-0778.
E-mail address: william.a.young1@ohio.edu.

are subjected to at the NFL Combine are outlined in Table 1 and will be discussed further later.

Table 1: Summary of Combine Measurements

Event Name	Measured Intent	Measured Units
Height	Physical Height	Inches
Weight	Physical Weight	Pounds
40-Yard Dash	Speed & Power	Seconds
20-Yard Dash	Acceleration & Explosiveness	Seconds
10 Yard Dash	Explosiveness & Quickness	Seconds
Shuttle	Speed, Explosion, & Change of Direction	Seconds
Cone	Agility, Change of Direction	Seconds
Vertical Jump	Lower Body Strength & Explosion	Inches
Broad Jump	Quickness & Lateral Burst Strength	Inches
225 Pound Bench Press	Upper Body Strength & Conditioning	Repetitions

Each of the events found in Table 1 translates to the skills required to play football in some fashion. For example, many of the drills start a player in a 3-point stance, which occurs when a person has two feet and one hand on the ground. The preceding movement from this initial stance simulates how a player will come off the line-of-scrimmage on a given play. In some cases, the player might be required to change his path of motion, while having the power and balance to engage opposing forces (i.e. a pass rushing from defensive linemen). Another drill might simulate the movement when a player is free to move off the line-of-scrimmage, but needs to achieve top running speed as quickly as possible without stopping and starting from rest (i.e. a wide receiver running a vertical route).

The assumption used in measuring the Combine drills is that the effects that a player receives from other players on the field of play can be individualized. It is believed that a player's football playing ability is correlated to the performance in these individual drills. If a cornerback runs the fastest 40-yard dash at the Combine, the skills should translate into covering skills needed against wide receivers during a play. It is conceivable that if the measured skills translate to football playing ability, the players with better performance measurements will turn out to be better professionals. The Combine also provides insight to a player's work ethic, which is vital in any stage of football. Therefore, assessing the results of the Combine events can be used as an aid for evaluating talent.

The 40-yard dash is a test that attempts to measure an athlete's speed and power. A player begins this drill in a 3-point stance and runs 40 yards in a straight-line path at max effort. During this event, the athlete is clocked at 10, 20, and 40 yards. Both the 20- and 10-yard dashes attempt to measure a player's explosiveness, which is a common term in football that signifies one's ability to move from rest. Talent evaluators interpret the results of the Combine events differently for different positions. A certain evaluator might place more significance on the results of a 10-yard dash for offensive and defensive linemen than his 40-yard performance. The logic for this ideology is that linemen run significantly less than 40 yards on a given play. However, the talent evaluator might place more significance on a 40-yard dash for receivers and cornerbacks since; they are required to run longer distances.

The 20-yard short shuttle is a measurement of lateral speed, explosion, and ability to change one's direction. The test begins from a 3-point stance where a player is positioned over a starting line. When the drill begins, the player runs five yards to one side and bends over to touch a line. At this point, the player runs back 10 yards in the opposite direction and bends to touch a second line. The drill finishes by the player returning to the starting point.

The 3-cone drill is a measurement of agility, cutting ability, and body control. In this drill, three cones are set up five yards apart in a triangular or "L" shape. The player begins in a 3-point stance at the first cone and moves to the second cone. He then returns to the starting cone and runs past the first cone to the third cone. Once a player is at the third cone, the athlete sprints back to the first cone while passing around the second cone on his way.

A vertical jump is a measurement of leg explosion and power, which is determined with a device with many plastic flags, or markers, hanging from a vertically standing pole. Before a player jumps he is flatfooted and the bottom of the pole is adjusted to the height of the individual's fingertips when extended above his head. The goal of the athlete is to disturb as many plastic markers as possible. This event is particularly important to wide receivers and defensive backs.

Broad jumps, or horizontal jumps, are measured when a player is asked to jump as far as he can into a pit of sand. The distance is measured from the starting line to the heal position when a player lands. This drill is thought to be an important measure for positions that require lower body strength, quickness, and lateral burst, which includes offensive and defensive linemen, as well as running backs and linebackers.

In the 225-pound bench press, a player is asked to lift 255 pounds as many consecutive times as possible. This drill is available for all position types at the Combine; however, many athletes do not participate. This test is believed to be a good measure of upper body strength and conditioning and is particularly important to offensive and defensive linemen.

Several other tests are conducted at the Combine, which include other physical measurements, intelligence tests, joint movement and flexibility tests, and drug tests. There are also position specific drills that are designed to test skills necessary at each position. Though there are several other attributes that are collected, this research will only investigate the performance measurements as shown in Table 1.

With the importance of the Combine defined, it is surprising that little research has been conducted to study the effects that aging has on one's performance over time in these events. The lack of research could be a result of the lack of publicly available data. Teams most likely collect a wide variety of data to assess and monitor physical characteristics about their players. However, the physical measurements that might be taken from professional strength and conditioning programs vary from team-to-team and trainer-to-trainer. Since a large percentage of current NFL players have competed at the Combine, a unique opportunity exists to evaluate the effects of aging in the standardized events. However, since repeated measurements are not taken or are not available to the public, the evaluation of aging is difficult. Therefore, to measure this effect, a methodology will be presented that is based on the ability of an artificial neural network (ANN) to translate a rating system into Combine-performance values. Therefore, the problem with repeated measurements is overcome, since the ratings for a player is available over many years of their professional career. Translating the rating system is significant, because each Combine event is specific to the skills needed to play football and is a combination of all of the ratings that are available. Simply investigating the change in ratings over time is not sufficient.

The following sections of this report are as follows;

Section 2 is dedicated to a review of literature regarding physiological aging and ANNs.

Section 3 describes the proposed rating-system-translation methodology.

Section 4 disseminates the results of the methodology.

Finally, Section 5 concludes with a discussion about relevant research findings.

2. Literature Review

The section reviews relevant information in the study of physiological aging. The intent is to discuss the effects that aging has on the human body and performance in sports. A short section about ANN theory and training practices will also be presented.

2.1. Physiological Aging

There has been significant research dedicated to the study of the human body's inability to retard the effects of aging. Some of these studies include; fingernail growth rate (Orentreich, Markovskyi, & Vogelmer, 1979), size of muscle fibers (Larsson, Grimby, & Karlsson, 1979), muscle cell density (Lexell, Henriksson-Larsen, Wimblad, & Sjostron, 1983), DNA repair rate (Wei, et al., 1993), hormone levels (Hakkinen & Pakarinen, 1993), muscular strength (Meltzer, 1994), pulmonary ventilation (Pollock, et al., 1997), and muscle enzymes (Ojanen, Rauhala, & Hakkinen, 2007). These physiological studies are correlated to athletic performance changes. Some examples of sports related performance declines are discussed below.

Studying athletic performance began as early as 1925, when data was collected to investigate performances for various World Records (Hill, 1925). Several examples of independent research indicates that athletic performance in swimming (Bortz & Bortz, 1996) and running (Starkes, Weir, & Young, 2003) declines at a rate of 0.5% per year after peak performance has been reached. Though this value seems to reoccur in many other related disciplines, activities that rely on cognitive-motor skills, rather than physical skills, have lower rates of decline (i.e. 0.35% per year after peak) (Baker, Hortonb, Pearceb, & Deakinb, 2005). Higher rates of decline have been found in competitive indoor rowing events on older subjects (i.e. 50 to 74 years of age), where the decline rate was found to be 0.7% per year after peak (Seiler, Spirduso, & Martin, 1998).

Most all research shows linear, curvilinear, or exponential trends when modeling the effects of aging. For example, when evaluating the effects of aging on freely chosen walking speeds, linear trends were found (Himann, Cunningham, Rechnitzer, & Paterson, 1988). Curve-linear trends were found when investigating competitions in indoor rowing events (Seiler, Spirduso, & Martin, 1998). Finally, freestyle swimming performance noted exponential performance decreases (Tanaka & Seals, 1997).

One downfall in studies investigating aging is that sufficient data is not always available to investigate the full aging process (i.e. performances until peak, peak performance, and performance after peak). When sufficient data is available, researchers will determine the various rates of improvement, as well as decline, and the age in which peak performance occurs. One example of a longitudinal study investigated the United States Masters Swimming Championships and collected data from teenager competitors to athletes in their mid 80's (Peterson & Simmons, 2007). This research indicated that athletes improved their performances by 3% per decade until peak was reached at 30 years of age, and declined at a rate of 7% per decade thereafter. For long distance endurance running, peak age was found to be 35 (Dempsey & Seals, 1995). Studies investigating peak age for professional baseball pitchers and batters achieve peak performance at 26 and 28 years of age respectively (Fair, 2008).

In the review of functional capacity articles, it appears that there is strong evidence supporting the fact that aging effects in performance are both gender and task specific (Lundel, Metter, & Lynch, 1997). Though there are varying degrees of changes in sports performance due to aging, the models used are constantly linear, exponential, or polynomial based when a large sample range is collected.

2.2. Artificial Neural Networks

Mathematical models, called ANNs, have been designed to imitate the primitive cognitive functionality of the human brain. Initial efforts to develop a biologically inspired model have been dated back to the 1940s (McCulloch & Pitts, 1943). ANNs are data driven and are a subset of artificial intelligence, called machine learning. These models are often called "universal approximators," (Reed & Marks, 1998) since they have a history of mapping input to output space accurately for a wide variety of problems (Efraim, Jay, Liang, & McCarthy, 2001). The structure of an ANN consists of interconnecting processing elements. These elements attempt to behave like biologically inspired neurons, where signals are processed and passed through the network based on activation values. Typically, the functions in an ANN implement hyperbolic tangent, sigmoid, or linear transfer functions. ANNs consist of several layers of neurons connected through weights. The purpose of the weights is to associate how independent values are ultimately processed to map dependent ones. Thus, weights can either work as excitation or inhibition signals that propagate through the network.

The values of the weights in an ANN are found by a process known as training. Training is an experiential procedure based on trial-and-error. During this procedure, various network structures should be tested in order to find the optimal network topology for the system in question. There are several ANN topologies, which include multi-layered perceptron (MLP), generalized feed-forward (GFF), radial based functions (RBF), and support vector machines (SVM) (Haykin, 1999). Before the learning algorithm is instantiated, a network structure must be determined, which includes the number of hidden layers and the number of processing elements within these layers. In some cases, the network might consist of too many or too few neurons, which will prevent the network from "learning" properly. Therefore, several networks should be trained to determine the "best" network.

There are several other reasons why the experimental procedure is necessary. Other problems can arise from the training algorithms that are used to adjust the initial random weights of an ANN. For example, back-propagation with momentum (Rumelhart & McClelland, 1986) is one of the most popular learning algorithms. As with all training algorithms, the network changes the weights based on an evaluation of a cost function. Typically, the cost function is a mean squared error. Thus, the network adjusts the values until the cost function is near optimal. There is a possibility that either the weights are changed too aggressively or not aggressively, depending on the parameters set for the training algorithm (i.e. learning or momentum rates). More advanced training algorithms have been developed that are based on optimization strategies (Levenberg, 1944) (Marquardt, 1963). However, second order training methods often come at a heavier computational cost.

ANN knowledge discovery begins by pre-processing data, which includes treating missing values, randomizing, normalizing, and translating symbolic values into unary-encoded data. Finally, the data is partitioned into three sub-sets of data, which include training, cross-validation, and testing. An oftenoverlooked practice in ANN training is the use of cross-validation, which promotes generalization and prevents the ANN from over-fitting (Princip, Euliano, & Lefebvre, 2000). Once the proper procedures of ANN training have been preformed and a final network has been determined, the network can then be used for a variety of modeling practices (Young, Weckman, Thompson, & Brown, 2008).

3. Methodology

The following section describes the methodology used to evaluate the effects of aging in NFL athletes using performance-aging curves. One problem in using the Combine dataset is that this measurement is only recorded once in a professional's career. Therefore, in order to obtain repeated measurements needed to evaluate the effects of aging, an ANN is used to translate a rating database that has been acquired. The following section is dedicated to explaining the databases that were obtained, ANN training, and the performance aging curve model.

3.1. Combine Database Acquisition

The NFL Combine data was obtained from a publishing company called the Sports Xchange (Sports Xchange, 2007). This company supplies NFL scouts, coaches, trainers, agents, and other analysts with specific results of the Combine events as well as their own analysis of draft eligible players. Due to proprietary issues, the information collected from this source will not be presented. However, a summary table of the data collected is presented in Table 1, where the average, standard deviation, and numeric counts are presented for an aggregated set of playing positions.

Table 2: Draft Combine 1996 - 2006 Summary

Pos.	40 Yd. Dash	Cone	Shuttle	Horizontal Jump	Vertical Jump	Bench Press
WR+CB	4.5±0.1(196)	6.9±0.2(126)	4.1±0.1(131)	123.1±5.6(136)	37.3±2.8(150)	15.1±4(80)
RB+FB+TE	4.6±0.1(126)	7.2±0.3(84)	4.2±0.2(85)	116.2±6(93)	34.5±2.7(96)	21.2±4.2(90)
OLB+MLB	4.7±0.1(133)	7.1±0.2(94)	4.2±0.2(96)	117.8±6.5(99)	35.3±3.5(102)	23.3±4.8(99)
SS+FS	4.5±0.1(79)	7±0.2(56)	4.1±0.1(54)	123±5.3(64)	37.6±3(64)	16.5±3.8(64)
OT+OG+C	5.2±0.2(185)	7.8±0.4(139)	4.7±0.2(140)	102.9±6.5(140)	29.6±3(144)	25.6±4.7(140)
DE+DT	4.9±0.2(133)	7.5±0.4(85)	4.5±0.2(90)	111.4±7.4(100)	32.7±3.3(105)	25.9±6.1(104)
QB	4.8±0.2(58)	7.2±0.3(37)	4.3±0.2(37)	111.2±7.7(37)	32.3±3.7(39)	N/A
K+P	4.9±0.2(19)	N/A	N/A	N/A	N/A	N/A

Playing positions were grouped together to increase sample size, which in turn increases statistical validity when fitting an aging curve model to the forecasted values (discussed later). Another reason the players were grouped together was because similar positions were assumed to experience the same effects of aging. For example, offensive linemen (offensive tackles, guards, and centers) were grouped together because these playing positions have similar body types, weights, and risk of injury while training or playing. Other groups were created based on similar assumptions. Additional data collected that will be used for the performance aging curves is found in Table 3.

Table 3: Additional Draft Combine 1996 - 2006 Summary

Pos.	Age	Height	Weight
WR+CB	24±1.2(196)	72±2.2(196)	198±15.1(196)
RB+FB+TE	24±1.3(126)	73±3.1(126)	234±22.7(126)
OLB+MLB	24±1.4(133)	74±1.4(133)	242±12.7(133)
SS+FS	23±1.1(79)	73±1.7(79)	208±10.7(79)
OT+OG+C	24±1.3(185)	77±1.8(185)	313±17.9(185)
DE+DT	24±1.2(133)	76±1.5(133)	287±23.9(133)
QB	24±0.8(58)	75±1.7(58)	223±11.1(58)
K+P	24±1.3(19)	74±1.5(19)	210±17.7(19)

Participants that are invited to the Combine do not always participate in every event. Many of the top players in the league option out of the events, such as the 40-yard dash. These players typically perform the events later at their college or training facility, which is known as a pro-day workout. The results of the pro-day work out are merged into the Combine dataset when a combine value does not exist. The data obtained for the pro-day workouts were also acquired from The Sports Xchange.

3.2. Rating Database Acquisition

Software manufacturers spend a significant amount of resources to design products that artificially simulate real-world environments. Electronic Arts developed one of the first NFL computer simulations in 1989, called Madden NFL Football (Electronic Arts, 2007). Improvements have been made over the years of development. Currently, Electronic Arts is one of few companies that have contracts with the NFL Players Association, which gives exclusive rights to the game manufacturer. Ironically, one of the original pioneers in the simulation development was Frank Cooney, which is the founder of the company in which the Combine data was acquired.

Developers acquired expert knowledge from professionals like John Madden in order to create a system that would increase the realism of the simulation. The simulation uses several attributes to model the characteristics of the player. These attributes (among others) include *Speed*, *Strength*, *Agility*, *Jumping*, *Stamina*, and *Injury*. A rating value, which is a value between 0 and 100, is assigned for each of the attributes for a player. The ratings are skewed towards the maximum, and are obtained from consulting companies like the Sports Xchange. Like the Combine data, specific data cannot be published in this report; however, a summary of the collected data is shown in Table 2, where the average, standard deviation, and numeric counts are provided.

Table 4: Madden NFL 2003 - 2008 Summary

Pos.	Speed	Strength	Agility	Acceleration	Jump	Stamina	Injury
WR+CB	90±3.4(196)	52±7.8(196)	89±4(196)	91±3.5(196)	87±5(196)	87±5.8(196)	85±6(196)
RB+FB+TE	82±8.9(126)	67±5.7(126)	81±8.9(126)	84±8.3(126)	73±7.9(126)	84±5.9(126)	83±5.8(126)
OLB+MLB	78±5.2(133)	72±4.4(133)	76±6(133)	82±5.8(133)	69±8(133)	83±6(133)	84±5.3(133)
SS+FS	87±3(79)	59±5.9(79)	85±3.4(79)	88±3.1(79)	84±4.1(79)	87±5.2(79)	84±6(79)
OT+OG+C	50±6.5(185)	86±3.7(185)	50±6.6(185)	64±9.2(185)	35±12.7(185)	72±6.8(185)	83±5.8(185)
DE+DT	65±7.7(133)	79±7.8(133)	64±7.6(133)	74±8.5(133)	59±10.9(133)	75±7.2(133)	83±4.3(133)
QB	63±9.6(58)	55±6.1(58)	63±10(58)	65±9.2(58)	58±11.1(58)	87±5.3(58)	84±7.1(58)
K+P	52±8.4(19)	39±10.9(19)	50±7.9(19)	54±10.7(19)	44±12.6(19)	75±13.5(19)	76±14(19)

3.3. ANN Training

The objective of the ANN is to determine how the rating values found in Table 4 and the additional Combine attributes found in Table 3 can be used to model the Combine performance values found in Table 2. A database for the ANN was composed by matching a player's Combine performance value with his rating value for the exact year match. Once this database was processed, the dataset was partitioned into three subsets of data for ANN training (60%), cross-validation (15%), and testing (25%).

Table 5: ANN Training Results for 40-yard Dash

Training Data	Value
MSE	0.011
Min Abs Error	0.000
Max Abs Error	0.382
r	0.939

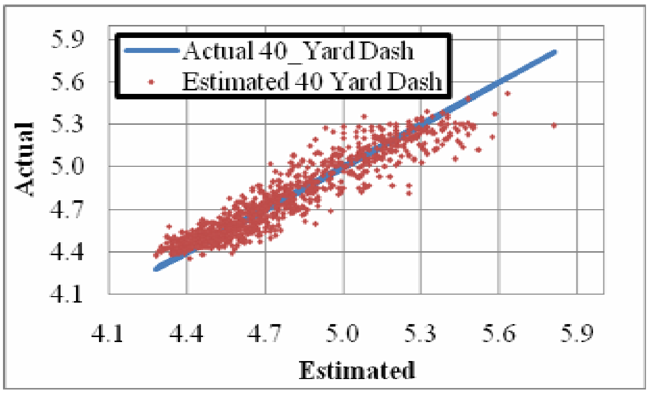

Fig. 1: ANN 1:1 Plot for 40-yard Dash Training

To demonstrate this methodology, an example of calculating a performance-aging curve for a 40-yard dash will be presented. To determine the "best" ANN for the rating system needed to predict the 40-yard dash times, various models were tested. This experiment included varying the network size, changing the transfer function types, trying different topologies, and using various training algorithms. The network that produced the best results was a two-hidden layer (1-5-4-1) network that was trained using conjugant gradient learning, and sigmoid processing elements. Further discussions about the experimental procedure are beyond the scope of this article; however, training results are shown below in Table 5.

A graphical representation of the network's accuracy used to predict the 40-yard dash times from the rating values is shown in Fig. 1, where the estimated values from the ANN are plotted along against 1:1 line, which is created by plotting actual verse actual values. This graph can be interpreted as estimates that are above the line are over estimated, and the values below the line are under estimated.

3.4. ANN Testing

To test the network, 25% of the original dataset was held out. The network performance values are shown in Table 6, where the quality of fit is adequate to use the ANN model for future forecasts. Once the network was validated, sensitivity about the mean analysis was performed. The result of this test quantifies the significance of input variables to the ANN model. For this test, the input values were varied from their sample mean by ±1 standard deviations with 50 steps per side of the mean. The values shown in Fig. 2 indicate that the two most significant input variables in the ANN model are *Weight*, and *Speed*. Thus, when these attributes are varied, they have the most influence in the model's output.

Table 6: Testing Results for 40-yard Dash

Performance	40_YARD
MSE	0.013
Min Abs Error	0.001
Max Abs Error	0.520
r	0.937

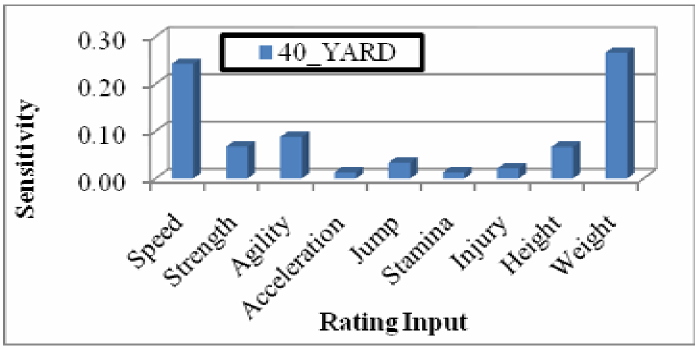

Fig. 2: Sensitivity Analysis for 40-yard Dash

The sensitivity analysis also is used to help explain the relationships used by the ANN when it formulates its estimate. For example, Fig. 3 and Fig. 4 show the change in the ANN output (40-yard Dash) when all other variables are held to their respective sample mean. A near-linear trend is observed when *Weight* is varied. This trend shows that as *Weight* increases, an "average" player's performance in the 40-yard Dash will decline (i.e. an increase in time). A non-linear trend is observed when the *Speed* rating is increased. In this trend, as *Speed* increases, there is a logical translation to the 40-yard Dash, where performance increase (i.e. time decreases).

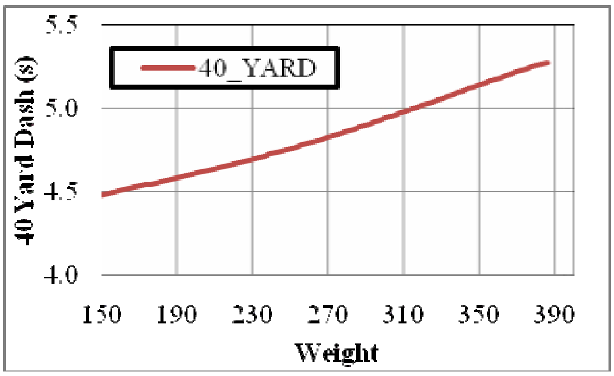

Fig. 3: Sensitivity for Weight

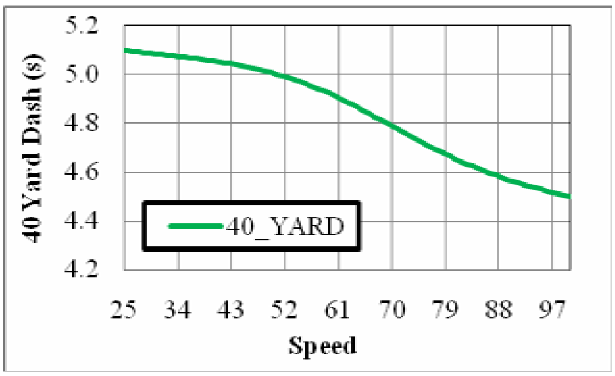

Fig. 4: Sensitivity for Speed

3.5. Aging Curve Model

The model used for the performance-aging curve is shown in Equation 1 (Young & Weckman, 2008), which is a modified polynomial equation. This standard form has been modified in order to make a forecast throughout a professional's career, the only variables that are required is Combine x_{age} and *Combine Value*$_{Xo}$

(i.e. 40-yard dash measurement at the Combine). Thus, once the factors (f_1, f_2, and f_3) of the model are determined and the two Combine values are obtained, a forecast can be made at any age (x_{age}) after the initial values were obtained.

$$Combine\,Value\pm_{x_{Age}} = Combine\,Value_{x_{age}} \pm t_{inv}(\alpha, df)\,|\,S\tan dard\ Error \quad (1)$$

Equation 1 is able to model three distinct periods of aging. The first period is when there is an improvement in athletic performance because of maturity and being exposed to a professional strength and conditioning program. The second period is when peak performance is reached. Finally, the third period is when there is a decline in athletic performance. The decline can potentially be explained by the accumulated trauma occurred from training or playing, or other aging related effects.

3.6. ANN Production

The entire rating database was used as a production (or forecasting) database even if there was not a match to be made with the Combine dataset. The two datasets (Combine and rating) have different ranges of years in which they were collected. Therefore, the small overlapping regions caused problems when formulating the parameters for the performance-aging model.

One method for determining an aging curve model with the available data would have been to match the initial Combine values with the forecasted Combine values from the ANN and apply a “family-based” effects model (Young, Masel, & Judd, 2008). However, this methodology was not used because the matching required would have resulted in a narrow range of age, which would not be sufficient to investigate the three periods of aging effects. Instead of investigating this short window of aging (i.e. 5 years), data aggregation was performed on the forecasted production dataset. This provided a much larger range (i.e. 25 years) of aging values, which made it possible to investigate all three aging periods. This type of data aggregation has also been used to reduce the effects of outliers (or noise) when constructing learning curves that evaluated efficiencies gained over time in manufacturing (Smunt & Watts, 2003).

An example of the data aggregation results is shown in Fig. 5. The values used in this figure were calculated by finding the conditional average of the actual and forecasted Combine values of the players’ age and position type. For this example, running backs, full backs, and tight ends were grouped together. To fit the data shown in Fig. 5 with Equation 1, *Average Combine*$_{X_0}$ and x_0 must be determined before the remaining factors (f_1, f_2, and f_3) are determined through statistical optimization. These values were found to be 4.62 ±0.15 seconds and 23.8±1.29 years respectively by averaging the data available.

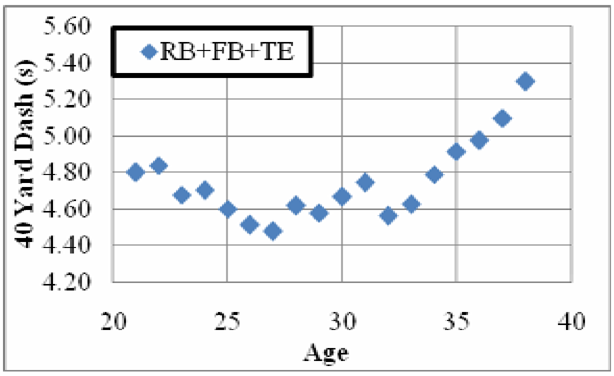

Fig. 5: Aggregated Aging Values

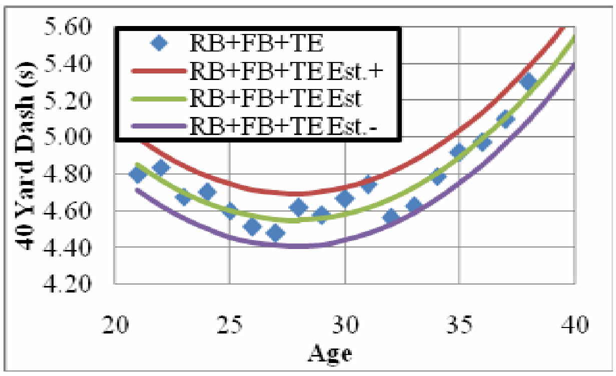

Fig. 6: Performance Aging Curve for 40 Yard Dash

Table 7: Aging Curve Model Parameters and Values for RB+FB+TE

Parameter	Estimate	Standard error	t-value df = 18	p-level	Lo. Conf Limit	Up. Conf Limit
f_1	1.01	0.01	184.64	0.00	1.00	1.02
f_2	-0.05	0.01	-6.22	0.00	-0.07	-0.04
f_3	0.01	0.00	9.55	0.00	0.01	0.01

Once these values are determined, the remaining factors in Equation 1 can be resolved for the values shown in Fig. 5. These factors are found by using an optimization algorithm (Levenberg-Marquardt), where parameters are determined by minimizing the predicted sum of square error. The factors and other relevant statistics determined for the performance aging curve model for the running back, full back, and tight end group is shown in Table 7.

The estimated factors and Combine values are used for the performance-aging curve shown in Fig. 6. The p-levels associated with the estimated parameters are all significant (i.e. p-value ≤ 0.05). When the factors were used form an estimate, the r-square value was 0.91. The estimates, as well as upper and lower confidence estimates are also shown in Fig. 6. The confidence intervals were calculated with Equation 2, where α, t_{inv}, and *Standard Error* was 0.05, 2.16, and 0.07 respectively.

$$Combine\,Value\pm_{x_{Age}} = Combine\,Value_{x_{age}} \pm t_{inv}(\alpha, df)\,|\,S\tan dard\;Error \quad (2)$$

4. Results

Further analyzing the performance-aging curve determined for the combined group of running backs, full backs, and tight ends for the 40-yard Dash produced the values found in Table 8. The peak age-performance age for this group was found to be 28 years of age, where an average Combine value of 4.62 seconds at 23.8 years of age was enhanced to 4.55 seconds. Thus, the average improvement rate found for this group is 6.11% until peak performance was met or an improvement of 0.87% per year. Once peak was reached, the loss encountered for this group was drastically different from the rate of improvement. An athlete from this group experienced a -2.69% decline in 40-yard performance per year.

Table 8: Aging Effects for RB+FB+TE

Aging Effect	Value
Peak Age	28
Peak Performance	4.55
%Gain to Peak	6.11%
%Loss After Peak	-48.35%
%Gain/year to Peak	0.87%
%Loss/year after Peak	-2.69%

Table 9: Aging Effects for Other Positions for the 40-yard Dash

Pos.	r-sq.	Avg. Combine ± Std. (n)	Avg. Combine Age ± Std. (n)	Peak Age	%Gain to Peak	%Loss after Peak	%Gain/ year to Peak	%Loss/ year after Peak
WR+CB	70%	4.5 ± 0.10 (21)	23.5 ± 1.22 (21)	28	0.15%	-1.24%	0.02%	-0.07%
RB+FB+TE	90%	4.6 ± 0.15 (18)	23.7 ± 1.29 (18)	28	6.11%	-48.35%	0.87%	-2.69%
OLB+MLB	87%	4.7 ± 0.12 (19)	23.8 ± 1.36 (19)	21	0.00%	-2.23%	N/A	-0.09%
SS+FS	72%	4.5 ± 0.08 (18)	23.5 ± 1.07 (18)	21	0.00%	-1.88%	N/A	-0.08%
OT+OG+C	66%	5.2 ± 0.18 (21)	24 ± 1.26 (21)	32	2.37%	-4.65%	0.22%	-0.33%
DE+DT	71%	4.9 ± 0.18 (20)	23.6 ± 1.25 (20)	26	1.89%	-33.60%	0.38%	-1.68%
QB	89%	4.8 ± 0.16 (21)	23.9 ± 0.85 (21)	21	0.00%	-6.18%	N/A	-0.25%
K+P	87%	4.9 ± 0.15 (24)	24 ± 1.29 (24)	29	1.32%	-7.07%	0.17%	-0.42%

Applying the rating-translation methodology and the performance-aging curve to the other positional groups for the 40-yard Dash resulted in the effects found in Table 9. The range of peak performance was found to be 21 to 32. However, some of these positional groups did not experience gains after the Combine (i.e. OLB+MLB, SS+FS, and QB). One possible reason why they did not improve might be that coaches or training staffs of these position groups do not train the athletes to maintain skills required for the 40-yard dash and develop other skills.

Other Combine events were examined in order to study the effects of aging. Some of these results are shown below in Fig. 7 through Fig. 10. Most of these curves show the shape of the polynomial equation; however, Equation 1 was able to model linear trends as shown in Fig. 7.

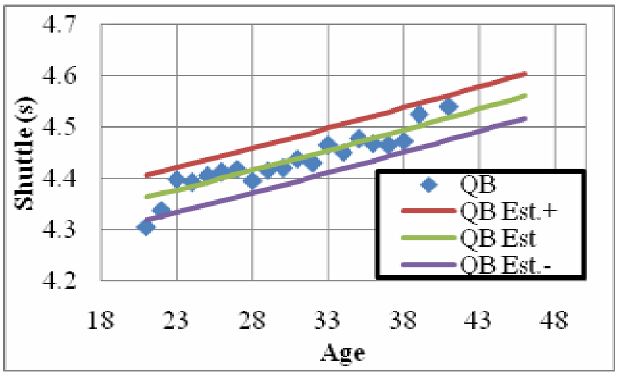

Fig. 7: Performance Aging Curve for Shuttle

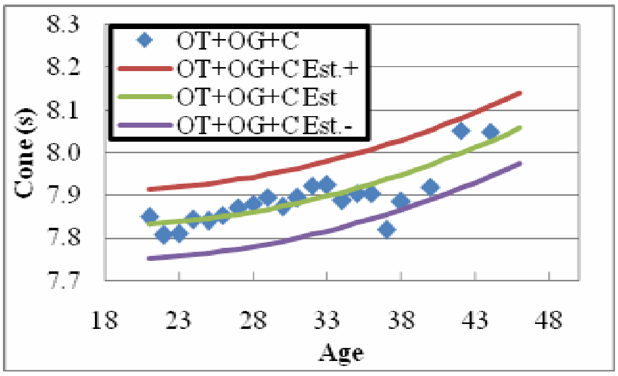

Fig. 8: Performance Aging Curve for Horizontal Jump

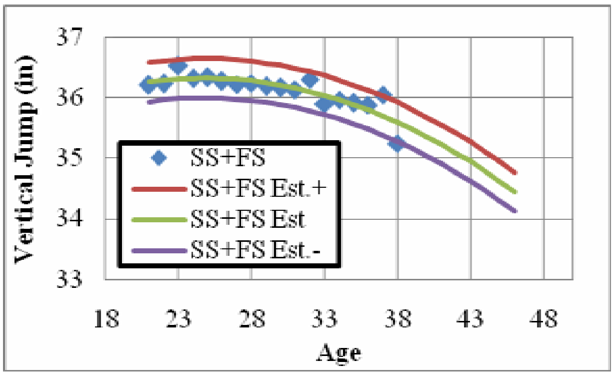

Fig. 9: Performance Aging Curve for Vertical Jump

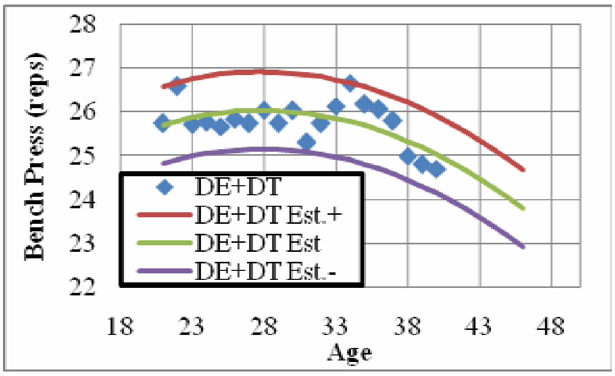

Fig. 10: Performance Aging Curve for Bench Press

Table 10 through Table 14 presents specific results of the effects of aging in other Combine events examined. In some instances, the aging performance models did not show adequate fits to the forecasted Combine values. Therefore, any positions that had an r-squared value less than 0.70 were removed from the tables.

Table 10: Aging Effects for Shuttle

Pos.	r-sq.	Avg. Combine ± Std. (n)	Avg. Combine Age ± Std. (n)	Peak Age	%Gain to Peak	%Loss after Peak	%Gain/ year to Peak	%Loss/ year after Peak
WR+CB	72%	4.1 ± 0.15 (21)	23.6 ± 1.22 (21)	29	0.90%	-4.46%	0.11%	-0.26%
RB+FB+TE	86%	4.2 ± 0.16 (18)	23.7 ± 1.18 (18)	21	0.00%	-10.93%	N/A	-0.44%
OLB+MLB	95%	4.2 ± 0.15 (19)	23.7 ± 1.21 (19)	21	0.00%	-16.38%	N/A	-0.66%
SS+FS	81%	4.1 ± 0.15 (16)	23.5 ± 1.02 (16)	21	0.00%	-1.12%	N/A	-0.04%
DE+DT	88%	4.5 ± 0.2 (20)	23.6 ± 1.3 (20)	25	0.56%	-14.40%	0.14%	-0.69%
QB	87%	4.3 ± 0.17 (20)	23.8 ± 0.83 (20)	21	0.00%	-4.55%	N/A	-0.18%

Table 11: Aging Effects for Cone

Pos.	r-sq.	Avg. Combine ± Std. (n)	Avg. Combine Age ± Std. (n)	Peak Age	%Gain to Peak	%Loss after Peak	%Gain/ year to Peak	%Loss/ year after Peak
OT+OG+C	74%	7.8 ± 0.39 (21)	23.9 ± 1.2 (21)	21	0.00%	-2.87%	N/A	-0.11%
RB+FB+TE	81%	7.2 ± 0.27 (18)	23.7 ± 1.17 (18)	21	0.00%	-7.27%	N/A	-0.29%
OLB+MLB	95%	7.1 ± 0.24 (17)	23.7 ± 1.21 (17)	21	0.00%	-2.91%	N/A	-0.12%
DE+DT	90%	7.5 ± 0.42 (20)	23.6 ± 1.32 (20)	24	0.29%	-13.52%	0.10%	-0.61%
QB	90%	7.2 ± 0.28 (20)	23.8 ± 0.83 (20)	21	0.00%	-3.80%	N/A	-0.15%

Table 12: Aging Effects for Bench Press

Pos.	r-sq.	Avg. Combine ± Std. (n)	Avg. Combine Age ± Std. (n)	Peak Age	%Gain to Peak	%Loss after Peak	%Gain/ year to Peak	%Loss/ year after Peak
OT+OG+C	86%	25.6 ± 4.66 (21)	24 ± 1.14 (21)	21	0.00%	-5.83%	N/A	-0.23%
DE+DT	70%	25.9 ± 6.06 (20)	23.6 ± 1.3 (20)	28	1.23%	-8.61%	0.18%	-0.48%
QB	81%	22 ± 3.98 (22)	24 ± 1.24 (22)	26	1.94%	-28.27%	N/A	-1.41%

Table 13: Aging Effects for Vertical Jump

Pos.	r-sq.	Avg. Combine ± Std. (n)	Avg. Combine Age ± Std. (n)	Peak Age	%Gain to Peak	%Loss after Peak	%Gain /year to Peak	%Loss/ year after Peak
RB+FB+TE	92%	34.5 ± 2.66 (17)	23.6 ± 1.2 (17)	21	0.00%	-13.46%	N/A	-0.54%
OLB+MLB	80%	35.3 ± 3.5 (19)	23.7 ± 1.19 (19)	21	0.00%	-0.87%	N/A	-0.03%
SS+FS	71%	37.6 ± 2.99 (18)	23.4 ± 1.04 (18)	25	0.17%	-5.15%	N/A	-0.25%
DE+DT	88%	32.7 ± 3.35 (20)	23.6 ± 1.28 (20)	24	0.23%	-13.43%	0.08%	-0.61%
QB	85%	32.3 ± 3.75 (22)	23.9 ± 0.83 (22)	21	0.00%	-7.57%	N/A	-0.30%

Table 14: Aging Effects for Horizontal Jump

Pos.	r-sq.	Avg. Combine ± Std. (n)	Avg. Combine Age ± Std. (n)	Peak Age	%Gain to Peak	%Loss after Peak	%Gain/ year to Peak	%Loss/ year after Peak
RB+FB+TE	90%	116.3 ± 6.03 (18)	23.6 ± 1.2 (18)	21	0.00%	-12.02%	N/A	-0.48%
OLB+MLB	75%	117.8 ± 6.46 (17)	23.7 ± 1.2 (17)	21	0.00%	-2.02%	N/A	-0.08%
SS+FS	70%	123 ± 5.27 (18)	23.4 ± 1.04 (18)	21	0.00%	-3.46%	N/A	-0.14%
DE+DT	88%	111.4 ± 7.42 (20)	23.6 ± 1.27 (20)	25	0.84%	-20.51%	0.21%	-0.98%
QB	89%	111.2 ± 7.71 (22)	23.8 ± 0.83 (22)	21	0.00%	-7.75%	N/A	-0.31%

5. Conclusions

Limited or censored data increases the difficulty of studying the effects of aging of professional football players in the NFL. Longitudinal data that is not proprietary is difficult to obtain; however, results from the NFL Scouting Combine is relatively easy to obtain through publishing companies like The Sports Xchange. However, Combine values are only available and recorded once during the career of an NFL player. Therefore, the proposed methodology takes advantage of a rating database developed by a computer software developer called Electronic Arts. The simulation software that has been developed rates players' physical qualities (i.e. *Speed*, *Agility*, and even their level of *Injury*). The developer of this system, which spends a significant amount of resources to make the simulation as accurate as possible, updates the rating values every year. When Combine values are paired with their respective rating values, a methodology can be created that investigates the effects of aging for NFL athletes using Combine performance values. This is accomplished by using an ANN to learn how the rating values represent actual Combine values. Once this has occurred, forecasts can be made using all of the ratings available.

Positional players were grouped together for the aging analysis, based on the assumption that players with similar playing positions aged similarly and would have similar bodily injuries risks that would affect their performances over time. The developed curves could be used to make forecasts for an expected Combine performance for a given player at any age, given that his initial Combine performance and age were known.

The performance aging curves that were presented in this research had the functionality to model three distinct periods of aging. When the performance-aging curve model parameters were determined for a grouped position type, it is assumed that the players would experience the same rates of improvement and decline. Using a player's Combine values in the aging curve allows players to age similarly, but account for the differences in initial performances recorded. Thus, given two players with the same age, the player who has a better Combine performance will have better forecasted values throughout his career.

The proposed methodology has the limitation in that the results are difficult to validate without expert knowledge in the field. However, the database translation using the ANN indicated good predicting performances as shown in Table 5 and Table 6. Thus, the legitimacy of the models ultimately relies upon the quality of the rating database. In an attempt NFL Head and Assistant Strength and Conditioning Coaches, who preferred not to be quoted were consulted. However, these discussions validated many of the aspects of this research anecdotally including; estimates for peak age, and improvement and decline rates.

Using professional athletes to isolate and model the effects of aging has many benefits. Professional athletes are extremely conditioned; thus, confounding problems with unhealthy or untrained individuals is nearly eliminated. However, due to the physical nature of professional football, players are subjected to brutally harsh conditions, where injuries occur at an extreme rate. Rates of aging performance decline are larger in professional football than other rates found in literature because of the implications of injury.

From the evaluation of the effects of aging in this research, it seems evident that there are different rates of aging for different Combine events. These rates are task specific, which is validated by past concepts found in literature. Each Combine event requires a unique set of skills to perform. The various task specific differences in rates might result from coaching staffs training their athletes differently now that they are at the professional level. This seems logical, since there are many differences between the collegiate and the professional game (i.e. size and speed differences).

Very few instances of aging research can be found involving professional football. However, some studies indicate strength is often used to discriminate starters against weaker (Barker, et al., 1993). Since younger inexperienced professionals typically do not start in the NFL, it seems logical that collegiate athletes improve in their physical performance over time. Therefore, it is likely that certain physical performance benchmarks must be maintained in order for a player to keep his starting position. One application of the research would be to investigate if certain benchmark physical conditions are heavily correlated with starting status, contract lengths or amounts, or other relevant studies in football.

Researchers could also use this methodology in many forms of applications in scientific reasoning. In other instances, sports evaluators might rate players in a similar manner, where players might have rated players in a non-statistical manner. For example, a particular scout might rate a running back with the following set of skills; “good” vision, “poor” run blocking, “above average” route running, and “great” ability to avoid tackles. For this example, a “fuzzy” rating system translation could be used to aid in executive decision making of hiring team members.

Finally, the presented methodology allowed an investigation of the effects of aging in the NFL. However, because of the way in which the data was prepared, the methodology may underestimate the effects of aging, because the data was analyzed for players who were able to stay in the league. These players, especially those who are older, must have been able to resist aging process and have not retired from playing. The older players examined possibly performed significantly better than others did in the Combine events. Thus, elite athletes with better initial performances values will be able to maintain their professional playing status longer. Athletes that were able to have long NFL careers could have biased the aging rates, since the sample size to aggregate becomes smaller for older players.

6. References

[1] J.Baker, S. Hortonb, W. Pearceb, J. Deakinb . A longitudinal examination of performance decline in champion golfers. *High Ability Studies* , 2005, **16** (2): 179–185.

[2] M. Barker, T. Wyatt, R. Johnson, M. Stone, H. O'Bryant, C. Poe, et al. Performance Factors, Psychological Assessment, Physical Characteristics, and Football Playing Ability. *Journal of Strength and Conditioning Research* , 1993, **7** (4): 224-233.

[3] W. Bortz , W. Bortz. How Fast Do Wwe Age? Exercise Performance Over Time as a Biomarker. *Journal of Gerontology: Medical Sciences* , 1996, **51**: 223–225.

[4] J. Dempsey, D. Seals. Aging, exercise, and cardiopulmonary cardiopulmonary. In D. Lamb, C. Gisolfi, & E. Nadel, *Medicine: Exercise in Older Adults,* 1995, **8**: 237–304. Carmel, IN: Cooper.

[5] T. Efraim, E. A. Jay, T. P. Liang , R. V. McCarthy. *Decision support systems and intelligent systems.* Upper Saddle River, NK: Prentice Hall. 2001

[6] Arts Electronic. *Madden NFL*. Retrieved January 24, 2007, from Madden NFL Homepage: *http://www.easports.com/*. 2007.

[7] R. Fair. Estimated Age Effects in Baseball. *Journal of Quantitative Analysis in Sports* , 2008,**4** (1),:1-39.

[8] K. Hakkinen, A. Pakarinen. Muscle strength and serum testosterone, cortisol and SHBG concentrations in middle-aged and elderly men and women. *Acta Physiol. Scand.* ,1993, **148**: 199–207.

[9] S. Haykin. *Neural Networks: A Comprehensive Foundation.* Prentice Hall.1999.

[10] A. Hill. The physiological basis of athletic records. *Lancet* , 1925, **209**: 483-486.

[11] J. Himann, D. Cunningham, P. Rechnitzer, D. Paterson. Age-related changes in speed of walking. *Medicine and Science in Sports and Exercis*e ,1988, **20**: 161-166.

[12] L. Larsson, G. Grimby, J. Karlsson. Muscle strength and speed of movement in relation to age and muscle morphology. *Journal Applied Physiology* , 1979, **46**: 451–456.

[13] K. Levenberg. A Method for the Solution of Certain Non-Linear Problems in Least Squares. *The Quarterly of Applied Mathematics* ,1944, **2**: 164–168.

[14] J. Lexell, K. Henriksson-Larsen, B. Wimblad, M. Sjostron. Distribution of different fiber types in human skeletal muscle: effect of aging studied in whole muscle cross sections. *Muscle & Nerve* ,1983, **6**: 588–595.

[15] R. Lundel,E. , Metter, N. Lynch. Age and gender comparisons of muscle.strength in 654 women and men aged 20-93 yr. *Journal of Applied Physiology* ,1979, **83**: 1581-1587.

[16] D. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*, 1963, **11**: 431–441.

[17] W. McCulloch, W. Pitts. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 1943, **5**: 115-133.

[18] D. Meltzer. Age dependance of Olympic weightlifting ability. *Medicine and Science in Sports and Exercise* , 1994, **26**: 1053-1067.

[19] National Invitational Camp. *The National Invitational Camp*. Retrieved January 20, 2007, from History: *http://www.nflcombine.net* . 2007.

[20] T. Ojanen, T. Rauhala, K. Hakkinen. Strength and power profiles of the lower and upper extremities in master throwers at different ages. *Journal of Strength and Conditioning Research* , 2007, **21** (1): 216–222.

[21] N. Orentreich, J. Markovskyi, J. Vogelmer. The effect of aging on the rate of linear nail growth. *Journal of Investigative Dermatology* , 1979, **73**: 126–130.

[22] M. Peterson, T. Simmons. Researcher's agine studies focus on entire lifespan. *OSU This Week*, 2007, **46** (20): 1-4.

[23] M. Pollock, L. Mengelkoch, J. Graves, D. Lowenthal, M. Limacher, C. Foster, et al. Twenty-year follow-up of aerobic power and body composition of older track athletes. *Journal of Applied Physiology*, 1997, **82**: 1508-1516.

[24] J. C. Princip, N. R. Euliano, W. C. Lefebvre. *Neural and adaptive systems: Fundamentals through simulations.* John Wiley and Sons. 2000.

[25] R. D. Reed, R. J. Marks. *Neural smithing: Supervised learning in feedforward artificial neural networks.* Cambridge, MA: MIT Press.1998.

[26] D. Rumelhart, J. McClelland. *Parallel Distributed Processing.* Cambridge, MA: MIT Press.1986.

[27] K. Seiler, W. Spirduso, J. Martin. Gender differences in rowing performance and power with aging. *Medicine and Science in Sports and Exercise* , 1998, **30**: 121-7.

[28] T. Smunt, C. Watts. Improving Operations Planning with Learning Curves: Overcoming the Pitfalls of Messy Shop Floor Data. *Journal of Operations Managment* , 2003, 93-107.

[29] Sports Xchange. *Premium NFL Draft Information*. Retrieved January 24, 2007 , from NFL Draft Scout: *http://www.nfldraftscout.com*. 2007.

[30] J. Starkes, P. Weir, B. Young. What does it take for older athletes to continue to excel? In J. Starkes, & K. Ericsson, *Expert performance in sports: advances in research on sport expertise*. Champaign, IL: Human Kinetics. 2003, **25**

[31] H. Tanaka, D. Seals. Age and gender interactions in physiological functional capacity: insight from swimming performance. *Journal of Applied Physiology* , 1997, **82**: 846-851.

[32] Q. Wei, G. Matanoski, E. Farmer, M. Hedayato, L. Grossman, et al. DNA repair and aging in basal cell carcinoma: a molecular epidemiologic study. *Proceedings of the National Academy of Sciences*, 1993, **90**: 1614–1618.

[33] W. Young, G. Weckman. Performance Aging Curves: Modeling the Effects of Aging through a Rating System. *Industrial Engineering Research Conference.* Vancouver, British Columbia. 2008.

[34] W. Young, D. Masel, R. Judd. A Matrix-Based Methodology for Determining a Part Family's Learning Rate. *Journal of Computers & Industrial Engineering* , 2008, **54**: 390-400.

[35] W. Young, G. Weckman, J. Thompson, M. Brown. Artificial Neural Networks for Knowledge Extraction of Concrete Shear Strength Prediction. *International Journal of Industrial Engineering* , 2008, **15** (1): 1-10.