FACULTY OF SCIENCES

Predicting Performance in Master Athletes

Word count: 8113

Nurefsan Davulcu
Student number: 01713636

Supervisor(s): Prof. Dr. Christophe Ley, Prof. Dr. Joern Rittweger

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Masters in Statistical Data Analysis

Academic year: 2020 – 2021

GHENT UNIVERSITY

Abstract

231 of the master athletes that competed in the 2018 World Masters Athletics Track and Field Championships (Malaga) were recruited by the German Aerospace Center (DLR) to complete a body composition assessment and two vertical jump tests, the countermovement jump test (CMJ) and hopping test. Multiple regression models of a subset of anthropometric, body composition, and vertical jump measures were able to explain 68.79%, 85.77%, and 60.65% of the variability in performance as a percent of the world record across runners and hurdlers, jumpers, and throwers, respectively. The performance of the running-hurdle model on a validation set resulted in a MSE of 0.5%. Also, the 5-fold cross-validation MSE was 0.7% and 0.6% for the jumping and throwing models, respectively. The CMJ peak power and the hopping test peak acceleration were significant predictors in runner-hurdlers, jumpers, and throwers ($p < 0.001$, for peak power and $p < 0.01$, $p < 0.01$, $p < 0.001$, for acceleration, respectively). The CMJ depth was a significant predictor in runner-hurdlers and throwers ($p < 0.05$ and $p < 0.01$, respectively). Body composition measures were not significant in any models. Finally, models involving measures from the jumping, hopping test, and/or body composition results are significantly better than a reduced model involving only simple measures height, weight, age, gender, and event. The results could be of use for future athletic assessment and placement of masters athletes.

Table of Contents

1 Introduction

1.1 Overview of Master Athletes

The formal structure of the master athletics discipline is relatively recent. Only in 1966, the first running event for men over 40 years was organized. In 1975, the first World Masters Championships were held where both men and women competed in running, jumping, and throwing events. In 1977, the international governing body named World Masters Athletics (WMA) was founded which now organizes the international competition biennially (1). The WMA Championships of 2018 (Malaga) counted over 8100 athletes from over 100 countries (2). There is no universal definition for master athletes as the definition varies widely among different sporting events (3). For the sporting event of interest in this study, it is defined as athletes 35 years and older (2).

Master athletes have been of interest because they are a positive example of successful aging. Older masters athletes have been shown to possess greater functional capacity at any given age than their sedentary peers. Three types of muscle deterioration with age have been discussed. Sarcopenia, describing an age-related decline in muscle mass; dynapenia, describing an age-related loss of strength and power; and sarcosthenia, describing an intrinsic weakening of skeletal muscle. Specifically, studying master athletes in comparison to their sedentary peers helps to understand the degree to which sarcopenia, dynapenia, and sarcosthenia are due to the direct effects of aging or a lack of physical activity (3).

1.2 Vertical Jump Tests

Vertical jump tests have been used in assessing athletic ability, in measuring the effectiveness of training programs, in talent identification, and in prediction of future success in specific athletic disciplines (4). A metric of high interest from a vertical jump test is peak neuromuscular power (3).

Different vertical jump tests exist including the squat jump (SJ), drop jump (DJ), the countermovement jump (CMJ), and a stiff leg jump (SLJ), also called a hopping test (5). The two jump tests of interest in this study are the CMJ and the 1-leg hopping test. In the case of the CMJ, a participant will start in a squat position on two feet and is instructed to jump vertically to their fullest potential landing in a squat position. The CMJ can be performed with hands on the hips or to swing freely. For a hopping test (1-leg or 2-leg variations), a participant will jump vertically starting from a stiff knee (standing) position (6,7).

1.3 Previous modelling of performance

The CMJ, hopping test, and other jump tests have been previously associated with and shown as strong predictors for performance across track and field events. For example, (5) investigated if variables from a variety of vertical jump tests could predict sprint performances for young sprinters through a regression analysis. The CMJ peak force relative to body weight significantly predicted maximal running velocity through 10-m and 60-m time. Furthermore, a planned comparison between sprinters and throwers showed significant differences in some vertical jump measures. Similarly, (8) conducted a controlled study on 33 basketball players that performed a CMJ test and sprinting over 10, 20, and 40m distances. CMJ height and maximum power adjusted for body mass showed statistically significant correlations with sprint performance times ($p<0.05$). These relationships were not shown when the CMJ measures were not adjusted for body mass (8). In addition, a study on 9 male shot-put athletes found significant correlations between the CMJ peak power ($r = 0.66$, $p<0.05$) and velocity ($r=0.70$, $p<0.05$) during competition period (9). Furthermore, a study that tracked the athletic performance of 24 university male throwers (shotput, discus, hammer, and javelin) for 2 years showed significant positive correlations across five different jump tests. The CMJ maximum height had an association of $r = 0.538$ ($p<0.001$) with performance (10). Finally, CMJ has been shown to be a useful warm-up method to increase performance in experienced hammer throwers (11) and shot-put throwers (12). As for a result from the hopping test (2-leg), (13) showed that leg stiffness and hopping power were significantly correlated with the maximal velocity of a 40-m sprint completed by 11 subjects ($r = 0.68$, $p < 0.05$ and $r= 0.66$, $p < 0.05$, respectively).

Anthropometric and body composition measures have also been associated with performance across track and field events. For instance, (14) took antropometric and blood sample measures from 11 male athletes. Maximal oxygen consumption (Vo2 max) and fat mass explained 83.9% of the variability in a multiple regression model for 27km trail run performance times (14). Furthermore, a review on East African runners, a group that has been consistently dominating long distance running, suggested that anthropometric and body composition measures were likely the determinants of their superior performance. A suggestion was made to look at these measures as explanatory rather than only descriptive as in many research studies (15). Lastly, an analysis conducted on 1353 middle-aged women and 1771 middle-aged men showed that running competition performance was negatively associated with fat mass index (FMI) through a regression analysis. Further, that FMI was a better predictor than body mass index (BMI) which is calculated only from the height and weight (16).

The previous literature has targeted specific athletic events such as a 100-m sprint, or a specific group such as sprinters. Furthermore, the analysis methods mostly involve comparing group differences or finding statistically significant correlations. Therefore, there are conclusions on associations of the variables but not necessarily measures of predictive performance. In addition, the sample sizes are mostly small which results in low statistical power. Lastly, the studies are targeted to young elite athletes.

The goal of this analysis will be to construct a model based on anthropometric, body composition, and vertical jump measures across all masters track and field events. It will be contributive because it will give an overview of all the track and field events, target the older age group of master athletes, and evaluating predictive power and not strictly association.

2 Participants and Data Collection

The German Aerospace Center (DLR) is the national center for aerospace, energy, transport, security and digitalisation research of Germany. In addition to conducting its own research projects, DLR also acts as the German space agency (17). 231 of the master athletes that competed in the 2018 WMA Track and Field Championships (Malaga) were recruited by the DLR to complete a body composition assessment and two vertical jump tests (CMJ and hopping test). The competition results include 32 different events across the disciplines of short-distance and long-distance running, jumping, throwing and the combined events decathlon and heptathlon. Some athletes participated in multiple different events and/or the same events more than once. The data contained 810 competition performances with male athletes aged 58.5 ± 12.2 (64.5%) and female athletes aged 55.2 ± 11.9 (35.5%).

2.1 Basic Assessment

The participants were assessed with a brief interview to assess information on athletic specialization, training habits, and medical conditions. Next, height and weight were assessed using a stadiometer (Kern MPE250 K100HM, Balingen, Germany) to the nearest of 0.1 cm and 100 g, respectively (6).

2.2 Countermovement Jump Test

The participants were asked to perform three maximal vertical jumps, with the instruction to raise the body as high as possible in the air. Subjects were given 30 to 90 s between jumps according to their own preference. Arms were allowed to swing freely, as long as no exaggerated movements were performed that could lead to an unnatural increase of jump height. After each jump, the validity was assessed. The test was performed with a Leonardo ground reaction force platform (Novotec Medical, Pforzheim, Germany) with the integrated software in its version 4.4b01.35 (research addition). This software screens the ground reaction force signal for a period in which movement-related variation is reduced below a critical threshold, to then provide a starting signal to participants. Power is computed as the product of force and velocity, and the peak jump power is the greatest power observed during the push phase, normalized to body mass. The Leonardo software also automatically computes the height of the jump (Hmax) and the depth of the countermovement (Hmin) (6).

2.3 Hopping Test (1-leg)

The test was performed with the same Leonardo software as for the CMJ with the program setting ‘multiple 1-legged jump’. The participants performed a series of vertical hops on one-leg with a stiff ankle. They started with low hops, worked up to high hops, then back to low hops. They were instructed to keep their ground contact time as short as possible as if they were on a ‘hot plate’ and to jump only on the forefoot. The arms were not to be raised above chest height.

2.4 Body composition Assessment

The body composition assessment was completed with a Bioelectrical Impedance Analyzer (Inbody S10, South Korea) following standard practices. All participants were asked not to take any foods or drinks and to avoid strenuous activity within 4 h before the testing. All parameters of BIA testing were obtained after removal of any metal parts from the body, and using a standard montage of outer and inner electrodes on the right hand and foot while patients lay down during 10 min before measurement with legs apart. Whole-body resistance (*R*) and reactance *(Xc)* were obtained by BIA using a single frequency, phase-sensitive 50 kHz. From the physical measurement, the integrated Inbody software computed the body composition measures including skeletal muscle mass (*SMM*), percent skeletal muscle mass (*PSSM*), soft lean mass (*SLM*), percentage of body fat (*PBF*), fat free mass (*FFM*), fat mass (*FM*), percent FFM (*PFFM*), percent FM (*PFM*), intracellular water (*ICW*), extracellular water (*ECW*), total body water (*TBW*), mineral content, and protein content.

Phase angle was calculated using:

$$phase\ angle\ (°) = arctan(reactance/resistance) \times (180/\pi).$$

Using resistance and reactance measured by BIA, apendicular skeletal muscle mass (ASMM) was estimated using the Sergi’s equation (18):

$$ASMM(kg) = -3.964 + (0.227 * RI) + (0.095 * weight) \\ +(1.384 * sex) + (0.064 * Xc)$$

Here, *RI* is resistance normalized for stature. Appendicular muscle mass index (ASMI, kg/m^2) was calculated by dividing ASMM by height in squared meter.

3 Preliminary Analysis

3.1 Software

All analyses were completed in R.

3.2 Clustered Data

Some athletes participated in the same event multiple times, making the data not iid. As this will be an issue for most modelling procedures, the data was de-clustered by choosing the best performance per participant per event. It was decided to choose the best performance because an athlete may deliberately not perform to their fullest capacity if they plan to perform in the event again, and further there may be an effect of fatigue in later events. A measure like average may penalize the athletes too much of their full performance capabilities. The delustering eliminated 144 observations in the full dataset resulting in 666 observations for analysis.

3.3 Further elimination

The decathlon, heptathlon, and throwing pentathlon scores will not be used in the modelling due to already having the performance that was used to calculate these scores in the dataset (n = 29).

3.4 Variable Selection

The correlation of the predictors was an important consideration. Most of the high correlations were logical, and due to either being calculated from the same physical measure or being a simple transformation. An example of a simple transformation was ASSM/height^2. Examples from measures that were modelled from the same physical measure occurred in the body composition, as described earlier in the methods, and vertical jump measures. Majority of the body composition assessment measures had $r > 0.7$ with each other, and with height and weight. Furthermore, jump height is dependent on countermovement depth in the CMJ (19) and reflected in the analysis ($r = -0.54$). Peak power and jump height were correlated with each other ($r = 0.92$) and correlated with phase angle ($r > 0.75$). For the hopping test, peak acceleration and contact time were correlated ($r = -0.74$).

A subset of predictors was chosen considering the statistical correlations discussed and physiological relevance and interest. The choice was also guided based on representation of all areas, that is, inclusion of measures from both jump tests and the body composition metrics. Body fat mass (*BFM*) and phase angle (*PhA*) were chosen among the body composition indicators, peak power (*PmaxTotRelCMJ*) and depth of the countermovement (*hminCMJ*) were chosen among the CMJ test, and peak acceleration (*Ftotmaxg_Hop*) was chosen among the hopping test. This was in addition to height, weight, age, gender, and event. The raw variables rather than transformations were used as some of the measures involved in the transformations, such as weight, were also included as predictors in the models.

3.5 Missing Data

The analysis of missing data on the de-clustered dataset showed that 2.3% of the bioimpedance measures and 10.7% of the hopping test measures were missing due to non-participation e.g., for having knee problems. The classification of the missing bioimpedance measures is unknown. The missing data for the hopping test can maybe be classified as missing not at random (MNAR) (20). Namely, a participant that may score worse on the test due to a knee problem is more likely not to participate. It was decided that listwise deletion would be the missing data approach due to the small amount of missing data and to not make further assumptions regarding the missingness.

3.6 Common Response variable: Age-Grading Performance and Percent of World Record

As the units of the performance across events differs, a common response variable was needed for unification. Upon initial discussion, the age grading performance constructed by Alan Jones in 2002, and updated the most recently in 2020, seemed a plausible response variable (21). The age grading performance adjusts the performance of a master athlete according to their age and gender, so that their performance is as if they were at the age of 27 (22). This idea was originally introduced in 1989 by the WMA through a booklet of age-graded tables to fairly assess masters athletic performance (21). The recent age factors are constructed from a piecewise function with quadratic and linear components as shown below, making the factors over age overall non-linear as illustrated in Appendix A.

$$f(x) = \begin{cases} 1 - C(c-b)^2 - B(b-x) - A(a-x)^2 & x \leq a \\ 1 - C(c-b)^2 - B(b-x) & a \leq x \leq b \\ 1 - C(c-x)^2 & b \leq x \leq c \\ 1 & c \leq x \leq d \\ 1 - D(x-d)^2 & d \leq x \leq e \\ 1 - D(e-d)^2 - E(x-e) & e \leq x \leq f \\ 1 - D(e-d)^2 - E(x-e) - F(x-f)^2 & x \geq f \end{cases}$$

Here, x represents the age, $f(x)$ represents the age-graded factor, a-e and A-E are parameters unique for each event and gender. The parameters *a-e* and *A-E* are estimated individually for each event and gender by fitting the function to the world records by age as illustrated in

Appendix A. The world records are then scaled to an interval [0,1] which represents the age-grade factors. The factored performance is then expressed as a percentage of the world record for that event (22).

$$age\ grade\ performance\ (\%)\ = \frac{performance\ *\ age - grade\ factor}{overall\ world\ record}$$

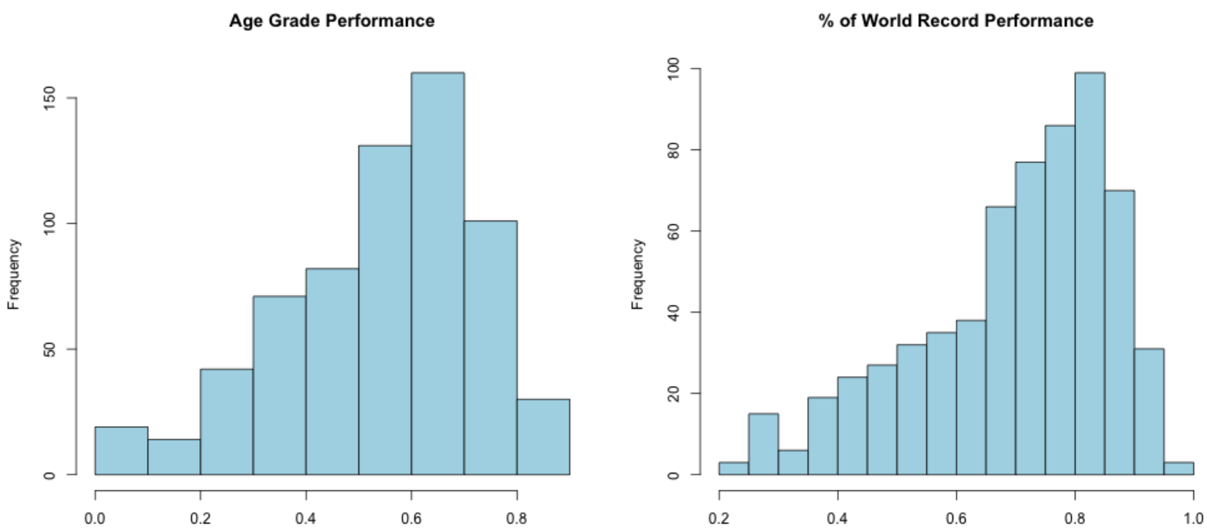


Figure 1: Distribution of Potential Response Variables: Age Grade Performance (left) and Percent of World Record Performance (right)

The distribution of the age-grading performance across all events was overall skewed as shown in Figure 1. For a simpler construction, the performance was simply expressed as a percent of world record per event, and age and gender would be included as predictors. The distribution of the percent of world record was comparable to the age-grading performance in shape, except the age-grading performance was centered at a higher percentage (better performance). This behaviour is expected since the age grade factors will upscale the performance of older athletes for fair comparison to younger athletes as described (Figure 1). However, the distribution of the two responses differed when visualized by event categories as shown in Figure 3.

4 Modelling Methods

4.1 Stratified Random Sampling

Stratified sampling involves random sampling from homogeneous, mutually exclusive sub-groups of the population called strata. The strata are formed based on common characteristics such as gender or regions within a country. It is effective at representing the entire population being studied compared to simple random sampling which is unbiased, but not always representative. Stratified sampling can be performed disproportionately or proportionately relative to the population. When performed proportionately, the strata reflect the same proportion in the sample as they do in the population. The sample size in a stratum is then calculated by (23):

$$n_s = \frac{total\ sample\ size\ *\ stratum\ size}{population\ size}$$

When performed disproportionately, the stratum sample size is simply:

$$n_s = \frac{total\ sample\ size}{number\ of\ strata}$$

4.2 Generalized Linear Models

A generalized linear model (GLM) extends on the classic linear regression model. It can be useful in cases where the assumptions of a linear regression are not fulfilled. A GLM has three components: the random component, the systematic component, and the link function (24).

a) *The random component* describes the distribution of Y, specifically the error components. In a linear regression model, we assume that

$$\varepsilon \sim N(0, \sigma^2)$$

In a glm context, we assume that the error terms can be distributed as any member of the *exponential family:*

$$f_Y(y;\theta;\emptyset) = exp\{\frac{y(\theta) - b(\theta)}{a(\varphi) + c(y,\varphi)}\}$$

For some specific functions $a(\varphi), b(\theta), c(y,\varphi)$

b) *The systematic component* describes the linear combination of the covariates that can be assigned to η

$$\eta = X\beta$$

where X is the matrix of covariates, and β is the matrix of model coefficients.

c) *The link function* describes the link between the random and systematic component. In a linear regression, the link function $g(\mu)$ is the identity function.

$$\eta = g(\mu)$$

4.3 Box-Cox Transformation

Box-cox is a power transformation used for asymmetrical distributions. The original proposal (25) is of the form:

$$y(\lambda) = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda}, & if\ \lambda \neq 0 \\ logy, & if\ \lambda = 0 \end{cases}$$

Maximum likelihood estimation is commonly used to determine which choice of λ succeeds best at transforming the data into a normal distribution (26). The log-likelihood that is maximized over (β, σ^2) for λ alone, referred to as the profile log-likelihood, is given by (27):

$$l_P(\lambda) = C - \frac{n}{2} log(s_\lambda^2)$$

$$\text{where } C = \frac{1}{\sigma^2} - \frac{n}{2} ln(2\pi)$$

$$s_\lambda^2 = \frac{y(\lambda, g)^T (I_n - G) y(\lambda, g)}{n}$$

$$y(\lambda, g) \sim N(X\beta, \sigma^2 I_n)$$

$$G = X(X^T X)^{-1} X^T$$

4.4 Variance Inflation Factors (VIF)

The VIF (28) is a measure of collinearity when performing regression analysis. It is the quantity $1/1 - \mathrm{R}_j^2$ in the variance of $\widehat{\beta_j}$ when p > 2 (29):

$$\mathrm{Var}(\widehat{\beta_j}) = \frac{\sigma^2}{1 - \mathrm{R}_j^2} \frac{1}{\sum_{i=1}^{n} (x_{ij} - \bar{x_j})^2}$$

where R_j^2 is the multiple R^2 for the regression of X_j on the other covariates. Further, the generalized VIF (30) used for coefficients with degrees of freedom greater than 1 is expressed as:

$$GVIF = VIF^{1/2df}$$

4.5 Subset Selection

One possibility of subset selection in regression is using stepwise procedures. Best subset regression finds for each k ∈ {0, 1, 2, . . . , p} the subset of size k that gives smallest residual sum of squares. Two algorithms which are computationally less intensive are forward and backward selection. Forward stepwise selection starts with the intercept, and then sequentially adds into the model the predictor that most improves the fit. Backward stepwise selection starts with the full model, and then sequentially removes predictors if it improves the fit. The classical way is to use the significance levels from an F-test to eliminate or keep predictors. However, an improvement of fit can be decided based on other criteria that are discussed in 4.6.

4.6 Modelling Selection and Validation Procedures

4.6.1 Overview

The best splitting of the data is given by splitting it into three components: training, validation, and a test data set. The validation set is used for model selection, that is, selection of the best model among others. The test data is used for model assessment, that is, assessing the prediction error on new data having chosen a final model. However, alternative methods exist for model selection when there is an insufficient amount of data to split it into three components. Two of these methods are usage of the cross validation and the AIC-criterion (31).

Prediction error has several variants two common being the mean-squared error (MSE) and the mean absolute error (MAE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i|$$

4.6.2 Adjusted R^2

The coefficient of determination (R^2) is not a good selection metric when comparing models with different complexity since the value will always increase with added complexity. Adjusted-R^2 is a modification of R^2 that penalizes the measure for added complexity (29).

$$R_a^2 = 1 - \frac{SSE/(n-p-1)}{SST/(n-1)} = 1 - \frac{n-1}{n-p-1}(1-R^2)$$

4.6.3 Training and Test Error

The training error is the prediction error on the training data. The test error is the prediction error on the test data which is used for model assessment. As the model is already adapted to fit the training data, the training error is optimistic and is not a good estimate of the test error. Training error consistently decreases with model complexity and approaches zero if we increase the model complexity enough. Further, a model with zero training error is overfit to the training data and will typically generalize poorly (31).

4.6.4 *K*-fold Cross validated Error

K-fold cross validation is a common estimate of the out-of-sample prediction error. The process involves splitting the data into K equally sized parts. Numbering the parts $k = 1,2..K$, we assign the kth part as the validation data. The remaining K-1 parts combined form the training data for model fitting. This process is repeated for all $k = 1,2..K$ and the average of the K prediction errors are taken. Typical choices of K are 5 or 10 (31).

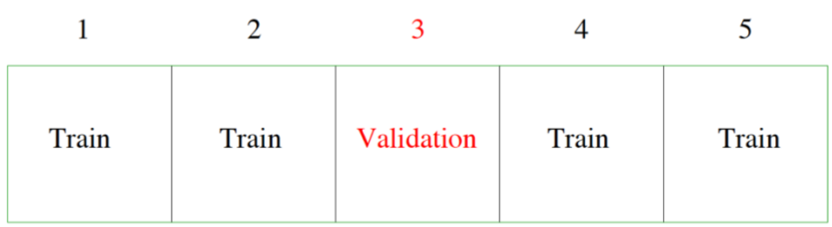


Figure 2: Illustration of 5-fold cross validation. k=3 is assigned as the validation data.

4.6.5 AIC Criterion

The AIC-criterion is an estimate of the in-sample prediction error (31). The likelihood function $\mathcal{L}(\hat{\theta} \mid y)$ is a measure of how likely one is to see their observed data given a model. The log-likelihood is used as a computational convenience. AIC uses a model's maximum likelihood estimation as a measure of fit, adding a penalty term for the complexity of the model. The log-likelihood is a measure of the bias while the penalty term $2k$ serves as a measure of the variance. Hence overall we have a balance of model fit with generalizability (32).

In the case of linear regression, the AIC is given by:

$$AIC = n \, log\left(\frac{RSS}{n}\right) + 2k, where\ RSS\ is\ the\ residual\ sum\ of\ squares \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

In the case of a generalized linear model, the AIC is given by:

$$AIC = -2\, log\mathcal{L}(\hat{\theta} \mid y) + 2k$$

The AIC value for a single model is not meaningful, rather it is meant to be used as a comparing metric between multiple models. A lower AIC value shows a better model (32).

4.6.6 General Linear F-test

The general linear F-test tests a full model with *p* parameters to a subset referred to as the reduced model with *k* parameters (33).

We can define the full model with *p* parameters to be:

$$Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon$$

And the reduced model with *k* parameters to be:

$$Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon$$

Now the hypothesis we want to test, where the null hypothesis is in favour of the reduced model is:

$$H_0: \beta_{k+1} = \beta_{k+2} = \ldots = \beta_p = 0$$
$$H_a: one\ of\ \beta_{k+1} = \beta_{k+2} = \ldots = \beta_p \neq 0$$

The full model will have *n-p-1* degrees of freedom and the reduced model will have *n-k-1* degrees of freedom. The F test statistic is then computed as:

$$F* = \frac{(SSE_{red} - SSE_{full})/(p-k)}{MSE_{full}} \sim F_{p-k,n-p-1}\ , if\ H_0\ is\ true$$

Where SSE_{red} and SSE_{full} are the sum of squared errors of the reduced and full models, respectively, and MSE_{full} is the mean squared error of the full model.

5 Results

5.1 Modelling Overview

Initially it was thought that one unified model for all performances would be the approach, however Figure 3 demonstrates that the distributions of running events, throwing events, and jumping events are distinct. Therefore, three separate models will be constructed. In addition, hurdle events are differentiated by the lower tail of the distribution from the running events however due to the small sample size (n = 62), it will be grouped with running events for the modelling. The listwise deletion will be applied separately for the running and hurdle data, jumping data, and throwing data for modelling to maximize the number of observations.

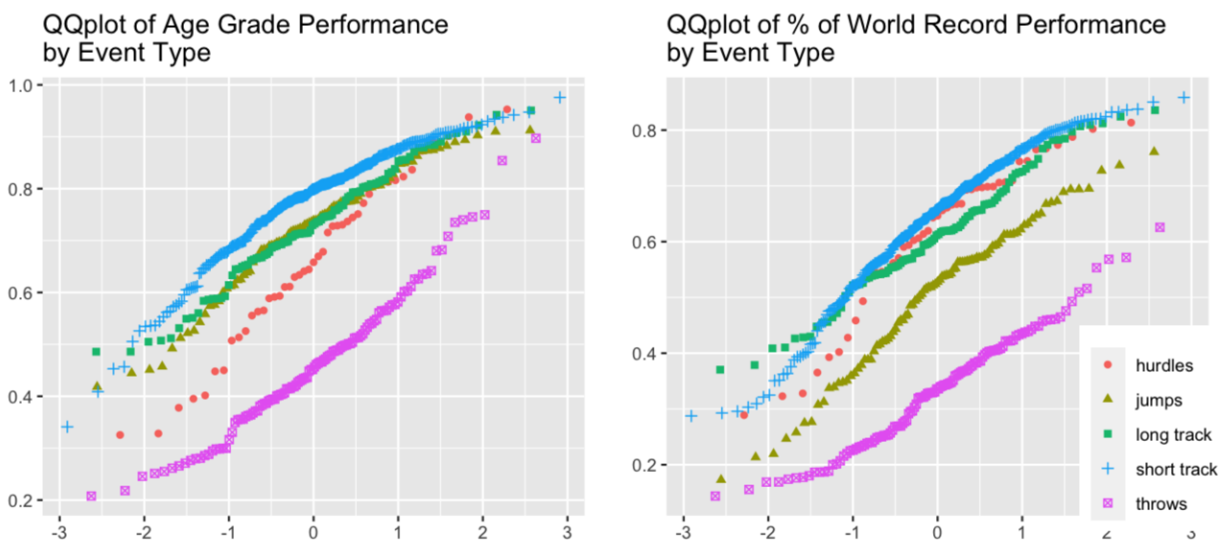


Figure 3: Distribution of Potential Response Variables by Event types. Age-graded performance (left) and Percent of World Record Performance (right).

Figure 4 shows the distribution of the response variable individually for the three groups. The throwing events showed a relatively uniform distribution outside of the tails which will be handled in the modelling steps.

The running and hurdle data was split into an 80/20 train and test dataset, respectively. Stratified sampling (4.1) by event was performed to ensure the proportion of data by event in the train and test datasets was equally represented. Further, a 5-fold cross validation (4.6.4), also sampling with stratification by event, was used on the training set for model selection. Since there was a small sample size for both throwing and jumping data (n<100), a test data set was not used. Instead, 5-fold cross validation was used for model assessment, that is, to estimate the test error.

Due to the small number of predictors and the relatively normal distribution of the response variables, a linear model seemed an appropriate method. A manual backward stepwise procedure was used starting from the full model, where at each step, the significance, standard errors, AIC, adjusted-R^2, model assumptions, and a 5-fold CV error or training error was assessed (4.6). A predictor was eliminated if it was not statistically significant however kept if the predictive performance of the model worsened as a general principle. Age and event were kept in any case, as controls.

From a practical point of view, Height, Weight, Age, Gender, and Event are variables that are easily attained compared to bioimpedance measures and results from the countermovement jumping and hopping tests. The model obtained after the selection procedure will be referred to as the final model. The model including only height, weight, age, gender, and event will be referred to as the reduced model. Each final model was compared using a general linear F-test (4.6.6) to the reduced model. AIC, adjusted-R^2, and prediction errors were also compared. Namely, to assess if the variables that are attained with higher resources are worthwhile to obtain for predicting performance.

Finally, the test error was calculated in the case of the running and hurdle model on the separate hold out set. For the throwing and jumping data, the models were fit on the entire dataset and the test errors were estimated using 5-fold cross validation, also stratifying by event, on the chosen models.

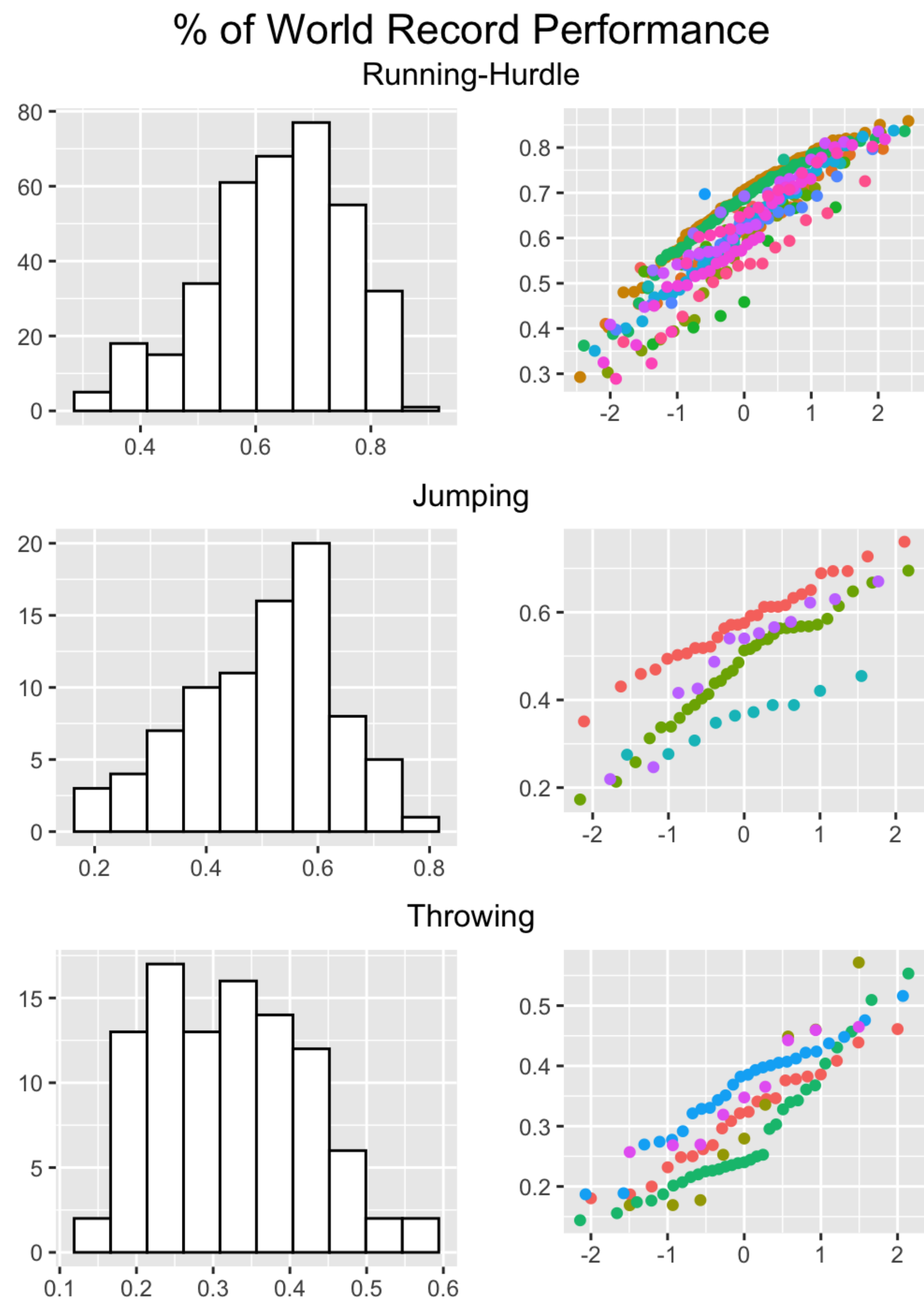


Figure 4: Distribution of Response Variable, Percent of World Record Performance by Model, Overall Histogram (left) and QQ plots by event (right)

Appendix B highlights the included events and Table 1 shows the means and standard deviations of each variable by model. Table 2 summarizes the modelling results for all three models: running and hurdle, throwing and jumping. The details for each model are provided in their respective sections.

Table 1: Descriptive Statistics for Model Variables (mean ± standard deviation)

	Running-Hurdle	**Jumping**	**Throwing**
Performance (% of World Record)	0.63 ± 0.12	0.5 ± 0.13	0.32 ± 0.1
Height (cm)	169.17 ± 8.19	171.97 ± 10.8	171.85 ± 9.67
Weight (kg)	67.3 ± 10.49	70.02 ± 11.59	76.1 ± 13.34
Age	56.75 ± 11.87	60.34 ± 14.06	60.25 ± 13.73
Peak Power CMJ (W/kg)	37.09 ± 9.16	37.04 ± 6.99	34.18 ± 7.25
Countermovement Depth CMJ (cm)	-0.27 ± 0.07	-0.26 ± 0.07	-0.25 ± 0.07
Peak Acceleration Hopping (g)	3.43 ± 0.43	3.38 ± 0.46	3.18 ± 0.38
Body Fat Mass (%)	13.38 ± 6	14.44 ± 7.41	19.1 ± 9.32
Whole body phase angle at 5kHz	6.1 ± 0.78	5.93 ± 0.73	5.89 ± 0.81

Figure 5: Correlations of Model Variables, record_performance is the response variable

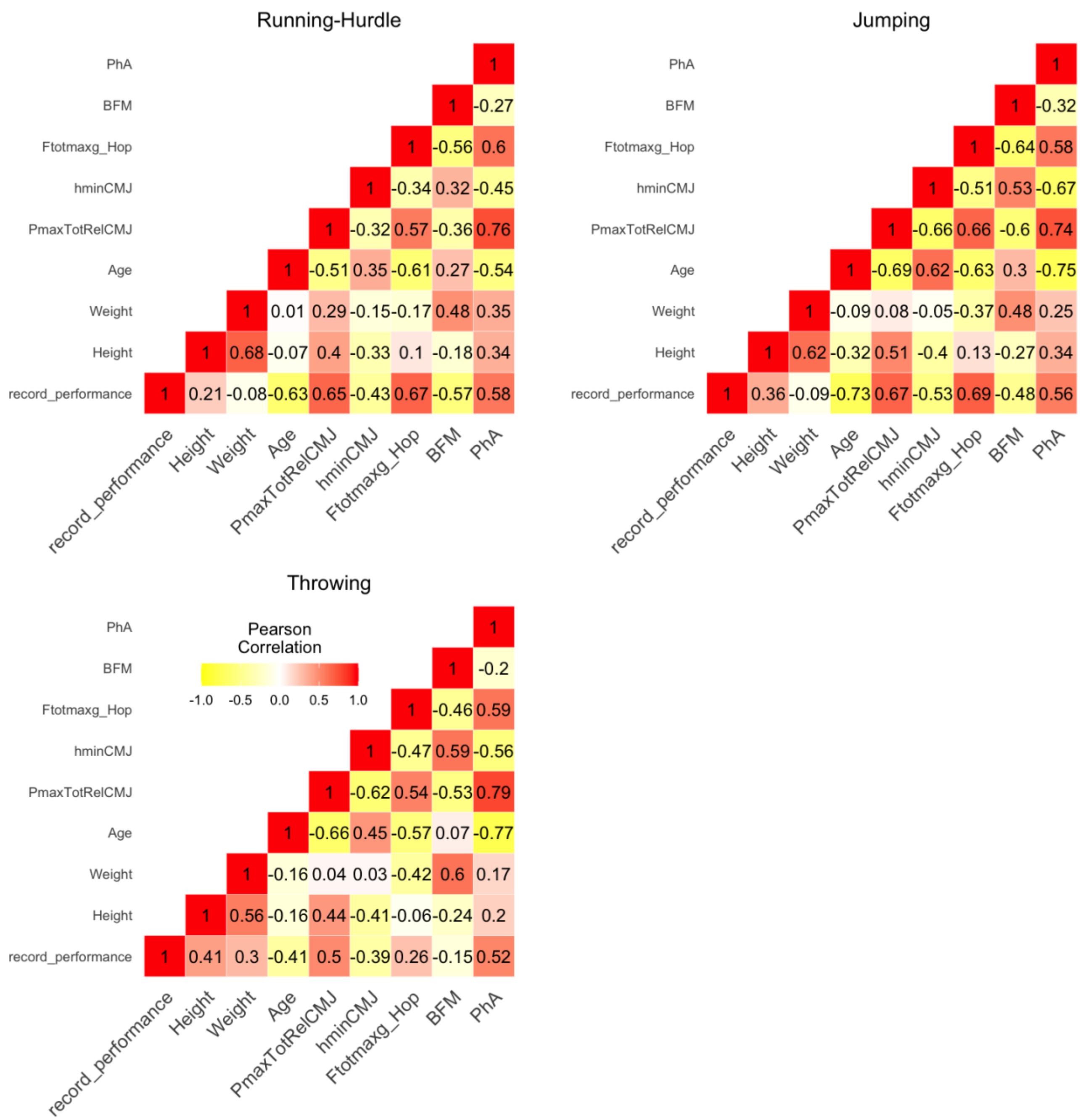


Table 2: Regression Model Summary by Model, [1]The response variable is on the log-scale

	Running-Hurdle	**Jumping**	**Throwing[1]**
N total for analysis	366	85	97
N (training)	286	none	none
N (5- fold CV)	55-59	16-18	19-21
N (test)	80	-	-
adjusted-R^2	68.79%	85.77%	60.65%
AIC	-708	-261	-20
Standard error	0.067	0.049	0.204
Training MSE	0.004	0.002	0.017
Training 5-fold CV MSE	0.004	-	-
Test Error or 5-fold CV MSE	0.005	0.007	0.006
Training MAE	0.047	0.036	0.108
Training 5-fold CV MAE	0.024	-	-
Test Error or 5-fold CV MAE	0.053	0.030	0.060

5.2 Running-Hurdle Model

There were 366 observations for running and hurdle events after listwise deletion and 1 observation removed for 200m hurdles. The 80/20 split resulted in 286 observations in the training data and 80 observations in the test data. The 5-fold split on the training set resulted in approximately equal groups of size ranging from 55-59.

From the full model, PhA was removed as a first step because it was insignificant ($p > 0.1$). Further, when the GVIF (4.4) were analyzed, weight and BFM were indicating collinearity as expected from the preliminary correlation analysis of the predictors. Removal of weight from the model resulted in a slightly better model than removal of BFM, however for the practical reason that weight is an easier measure to collect and because the improvement was small, BFM was removed for the final model. The resulting linear model for the running and hurdle data was then:

$$Y_i = Intercept + \sum_{j=1}^{17} Event_j * x_{ji} + Height * x_{18i} + Age * x_{19i} + Weight * x_{20i}$$
$$+ Gender * x_{21i} + \mathrm{PmaxTotRelCMJ} * x_{22i} + \mathrm{hminCMJ} * x_{23i} + \mathrm{Ftotmaxg_Hop} * x_{24i} + \varepsilon_i$$

The resulting adjusted-R^2 on the training data was 68.79% with a residual standard error of 0.067. The results of ANOVA showed Height, age, gender, weight, event, and PmaxTotRelCMJ were significant with $p < 0.001$. Ftotmaxg_Hop was significant with $p < 0.01$ and hminCMJ was significant with $p < 0.05$. There was no clear violation of homoscedasticity, normality, or linearity. Only a few outliers caused a small deviation from normality in the tails.

The full model was significant compared to the final model at the 5% significance level. However, the difference in adjusted-R^2 (~ -0.5%) and AIC-criterion from the full model to the final model was negligible giving support for the more parsimonious model. The final model was significant to the reduced model involving only height, weight, age, gender, and event ($p < 0.0001$). In addition, the final model shows an improvement in adjusted-R^2 (~ +5%), AIC-criterion, and standard error (-0.005) compared to the reduced model. Finally, the MSE from the 5-fold CV is the same for the full model, the final model, and the reduced model (0.004). The test error on the validation set was 0.005.

Table 3: Linear Regression Coefficients and 95% Confidence Intervals for Running-Hurdle Model.

*Significance levels (t-statistic) **** 0.001, *** 0.01, ** 0.05, * 0.1*

Variable	Coefficient	95% CI
(Intercept)	0.27**	[0.011, 0.528]
Height	0.003****	[0.002, 0.005]
Age	-0.004****	[-0.005, -0.002]
Weight	-0.004****	[-0.006, -0.003]
GenderM	0.039***	[0.013, 0.065]
Event10 km Road Race (run)	-0.077***	[-0.132, -0.021]
Event100m	-0.038	[-0.09, 0.014]
Event110m Hurdles	-0.134*	[-0.277, 0.008]
Event1500m	-0.099****	[-0.155, -0.043]
Event20 km Race Walk	-0.028	[-0.096, 0.041]
Event2000m Steeplechase	-0.1**	[-0.176, -0.023]
Event200m	-0.037	[-0.091, 0.016]
Event3000m Steeplechase	-0.112	[-0.256, 0.031]
Event300m Hurdles	-0.059	[-0.132, 0.015]
Event400m	-0.058**	[-0.113, -0.004]
Event400m Hurdles	-0.133*	[-0.275, 0.01]
Event5000m	-0.059**	[-0.118, 0]
Event5000m Race Walk	-0.011	[-0.086, 0.065]
Event8 km Cross Country	-0.047	[-0.104, 0.009]
Event800m	-0.074***	[-0.13, -0.019]
Event80m Hurdles	-0.054*	[-0.115, 0.007]
EventHalf Marathon	-0.129****	[-0.191, -0.067]
PmaxTotRelCMJ	0.003****	[0.002, 0.005]
hminCMJ	-0.179***	[-0.314, -0.044]
Ftotmaxg_Hop	0.043***	[0.014, 0.072]

5.3 Jumping Model

There were 85 observations for jumping events after listwise deletion. The 5-fold split for validation resulted in approximately equal groups of size ranging from 16-18.

BFM, PhA, hminCMJ, and gender were eliminated sequentially analyzing the model after each variable removal as described in the methods. The resulting linear model was:

$$Y_i = Intercept + \sum_{j=1}^{3} Event_j * x_{ji} + Height * x_{4i} + Age * x_{5i} + Weight * x_{6i} \cancel{+ Event * x_{7i}} + \text{PmaxTotRelCMJ} * x_{8i} + \text{Ftotmaxg\_Hop} * x_{9i} + \varepsilon_i$$

The resulting adjusted-R^2 on the full data was 85.77% with a residual standard error of 0.05. All variables except Ftotmaxg_Hop were significant with $p<0.001$, Ftotmaxg_Hop was significant with $p<0.01$.

There was some heteroscedascity indicated by the residual plots, attempts of common transformations of the response variable log, inverse, and square root were not successful. In addition, an attempt of a box-cox transformation (4.3) that maximized the likelihood function resulted in a λ value close to 1, indicating no transformation (25). An analysis of the residuals by the predictor variables showed the heteroscedastic effect in age. However, the transformation of age was not effective either.

The full model was not significant compared to the final model ($p > 0.4$), and there was no difference in standard error and a small improvement in adjusted-R^2 (< +1%) from the full model to the final model. The final model was significant to the reduced model involving only height, weight, age, gender, and event ($p < 0.0001$). There was an improvement in adjusted-R^2 (+3.8%), AIC-criterion, and standard error (-0.006) compared to the reduced model. The 5-fold CV MSE, serving as the estimate of the test error, was 0.007.

Table 4: Linear regression coefficients and 95% confidence intervals for Jumping Model

*Significance levels (t-statistic) **** 0.001, *** 0.01, ** 0.05, * 0.1*

Variable	Coefficient	95% CI
(Intercept)	0.15	[-0.117, 0.416]
Height	0.003***	[0.001, 0.004]
Age	-0.003****	[-0.005, -0.002]
Weight	-0.002***	[-0.004, -0.001]
EventLong Jump	-0.098****	[-0.123, -0.072]
EventPole Vault	-0.18****	[-0.217, -0.144]
EventTriple Jump	-0.069****	[-0.102, -0.036]
PmaxTotRelCMJ	0.003**	[0, 0.006]
Ftotmaxg_Hop	0.064***	[0.022, 0.105]

5.4 Throwing Model

There were 97 observations for throwing events after listwise deletion. The 5-fold split for validation resulted in approximately equal groups of size ranging from 19-21.

The AIC was not used as a comparing metric for the throwing model as some transformations were applied in the process and comparing of different dependent variables is not possible (32). PhA and BFM were removed sequentially from the full model. The adjusted-R^2 was almost unchanged. In the second step, removal of weight instead of BFM showed worse results. Further, removal of weight, then gender and vice versa, after BFM decreased the adjusted-R^2 by ~10%, so they were kept despite being not significant in the model. It can be noted that the GVIF (4.4) for age was ~8 so possibly some collinearity between weight and age was causing this result. Further eliminations yielded worse results as shown in Appendix C.

The model residuals showed heteroscedasticity and a violation of normality. The distribution of the response variable was relatively uniform outside of the tails. A log-transformation of the response was less uniform, however yielded a left-skewed distribution.

A linear model with a log transformed response variable and a gaussian generalized linear model (4.2) with a log-link function ($\eta = log(\mu)$) were fit to attempt to improve the violations. In the linear model, the log transformation improved both the heteroscedasticity and normality. However, the adjusted-R^2 dropped by ~5%. The glm log-link improved the normality and heteroscedasticity, although it was not as successful compared to the linear model as shown in Appendix D. However, the training error for the glm log-link was slightly lower than both the untransformed and the log transformation model. In the end, the log transformation linear model was chosen for the better improvement in model assumptions, and because the predictive performance was rather comparable to the glm log-link.

The resulting linear model can be written as:

$$log(Y_i) = Intercept + \sum_{j=1}^{4} Event_j * x_{ji} + Height * x_{5i} + Age * x_{6i} + Gender * x_{7i}$$
$$+ Weight * x_{8i} + \text{PmaxTotRelCMJ} * x_{9i} + \text{hminCMJ} * x_{10i} + \text{Ftotmaxg\_Hop} * x_{11i} + \varepsilon_i$$

The resulting adjusted-R^2 was 60.65% with a standard error of 0.204 and a mean-squared training error of 0.017. Height, age, weight, event, PmaxTotRelCMJ, and hminCMJ were significant with $p < 0.001$. Ftotmaxg_Hop was significant with $p < 0.01$.

The full model was not significant compared to the final model ($p > 0.6$). The adjusted-R^2 from the full model to the final model was negligible (<1%). The final model was significant compared to the log-transformed reduced model involving only height, weight, age, gender, and event ($p < 0.0001$). In addition, the final model shows an improvement in adjusted-R2 and the mean-squared training error (+21.86% and -0.002, respectively). The 5-fold CV MSE, serving as the estimate of the test error, was 0.006.

Table 5: Linear regression coefficients (log-scale) and 95% confidence intervals for Throwing Model

*Significance levels (t-statistic) **** 0.001, *** 0.01, ** 0.05, * 0.1*

Variable	Coefficient	95% CI
(Intercept)	-6.89****	[-8.559, -5.221]
Height	0.003	[-0.004, 0.01]
Age	0.02****	[0.011, 0.029]
GenderM	-0.452****	[-0.633, -0.27]
Weight	0.016****	[0.01, 0.022]
EventHammer Throw	0.018	[-0.149, 0.184]
EventJavelin Throw	-0.141**	[-0.258, -0.025]
EventShot Put	0.181***	[0.061, 0.301]
EventWeight Throw	0.232***	[0.067, 0.397]
PmaxTotRelCMJ	0.034****	[0.022, 0.046]
hminCMJ	-2.041****	[-3.099, -0.983]
Ftotmaxg_Hop	0.376****	[0.182, 0.57]

6 Discussion

Neither body composition measures, phase angle nor body fat mass, are present in the final models. The response variable and phase angle were moderately positively correlated for the running-hurdle, jumping, and throwing data ($r = 0.58$, 0.56, and 0.52, respectively). Similarly, for body fat mass and the response variable ($r = -0.57$, -0.47, -0.15, respectively). However, in both cases the correlation with the other predictors, likely caused them not being significant in the final model. Specifically, phase angle was strongly correlated with CMJ peak power present in all models. Similarly, body fat mass was strongly correlated with weight and present in all models.

One limitation of the models is the non-uniform distribution of events, which causes the models to be biased towards events with more data. This could be solved with a disproportionate stratified sampling rather than proportionate, namely sampling an equal amount of data per event, however this would simply eliminate too much data in this dataset. Furthermore, as different events are grouped together for a model, there is some loss of information. Specifically, the variation between events in the same category are not being captured. In the literature, the associations made have been targeting mostly specific events e.g. 100-m sprint. Specifically for the running and hurdle model, there is a big range of events covered. For instance, (34) compared characteristics between middle-distance and long-distance elite male young runners and found significant differences in their height, weight, relaxed upper arm girth, flexed and tensed upper arm girth, total upper arm area, upper arm muscle area, and thigh muscle area.

Another approach to the modelling could be to first perform a principal components analysis on the full set of predictors as a dimensionality reduction method. This would reduce the highly correlated body composition measures for example to one component. However, the interpretation of the individual variables could not be interpreted with this approach.

7 Conclusion

Anthropometric and vertical jump measures were good predictors of performance among runners, hurdlers, jumpers, and throwers in master athletes. The results could be of use for future athletic assessment and placement and contributive to the interest in masters athletics as a successful model of aging.

References

1. Olson LT. Masters Track And Field : A History. McFarland Publishing; 2000.

2. World Masters Athletics [Internet]. Available from: https://world-masters-athletics.com/

3. Tanaka H, Tarumi T. Aging and Physiological Lessons from Master Athletes. 2020;10(January):261–96.

4. Klavora P. Vertical-jump Tests : A Critical Review. Strength Cond J. 2000;22(5):70.

5. Markström JL, Olsson C-J. Countermovement Jump Peak Force Relative to Body Weight and Jump Height as Predictors for Sprint Running Performances. J Strength Cond Res [Internet]. 2013 Apr;27(4):944–53. Available from: https://journals.lww.com/00124278-201304000-00010

6. Alvero-Cruz JR, Brikis M, Chilibeck P, Frings-Meuthen P, Vico Guzmán JF, Mittag U, et al. Age-Related Decline in Vertical Jumping Performance in Masters Track and Field Athletes: Concomitant Influence of Body Composition. Front Physiol [Internet]. 2021 Apr 1;12(April):1–14. Available from: https://www.frontiersin.org/articles/10.3389/fphys.2021.643649/full

7. Veilleux LN, Rauch F. Reproducibility of jumping mechanography in healthy children and adults. J Musculoskelet Neuronal Interact. 2010;10(4):256–66.

8. Shalfawi SA, Sabbah A, Kailani G, Tønnessen E, Enoksen E. The Relationship Between Running Speed and Measures of Vertical Jump in Professional Basketball Players: A Field-Test Approach. J Strength Cond Res [Internet]. 2011 Nov;25(11):3088–92. Available from: https://journals.lww.com/00124278-201111000-00021

9. Kyriazis TA, Terzis G, Boudolos K, Georgiadis G. Muscular Power, Neuromuscular Activation, and Performance in Shot Put Athletes At Preseason and at Competition Period. J Strength Cond Res [Internet]. 2009 Sep;23(6):1773–9. Available from: https://journals.lww.com/00124278-200909000-00021

10. Takanashi Y. The relationship between jump ability and athletic performance in athletic throwers. Sport Mont. 2021;19(1):71–6.

11. Karampatsos G, Terzis G, Polychroniou C, Georgiadis G. Acute effects of jumping and sprinting on hammer throwing performance. J Phys Educ Sport. 2013;13(1):3–5.

12. Terzis G, Karampatsos G, Kyriazis T, Kavouras SA, Georgiadis G. Acute Effects of Countermovement Jumping and Sprinting on Shot Put Performance. J Strength Cond Res [Internet]. 2012 Mar;26(3):684–90. Available from: https://journals.lww.com/00124278-201203000-00013

13. Souhaiel MC, Denis C. Leg power and hopping stiffness: Relationship with sprint running performance. Med Sci Sports Exerc. 2001;33(2):326–33.
14. Alvero-Cruz JR, Parent Mathias V, Garcia Romero J, Carrillo de Albornoz-Gil M, Benítez-Porres J, Ordoñez FJ, et al. Prediction of Performance in a Short Trail Running Race: The Role of Body Composition. Front Physiol [Internet]. 2019 Oct 16;10. Available from: https://www.frontiersin.org/article/10.3389/fphys.2019.01306/full
15. Mooses M, Hackney AC. Anthropometrics and body composition in east african runners: Potential impact on performance. Int J Sports Physiol Perform. 2017;12(4):422–30.
16. Herrmann FR, Graf C, Karsegard VL, Mareschal J, Achamrah N, Delsoglio M, et al. Running performance in a timed city run and body composition: A cross-sectional study in more than 3000 runners. Nutrition. 2019;61:1–7.
17. DLR at a glance [Internet]. 2021. Available from: https://www.dlr.de/EN/Home/home_node.html
18. Sergi G, De Rui M, Veronese N, Bolzetta F, Berton L, Carraro S, et al. Assessing appendicular skeletal muscle mass with bioelectrical impedance analysis in free-living Caucasian older adults. Clin Nutr [Internet]. 2015 Aug;34(4):667–73. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0261561414001861
19. Salles AS, Baltzopoulos V, Rittweger J. Differential effects of countermovement magnitude and volitional effort on vertical jumping. Eur J Appl Physiol [Internet]. 2011 Mar 30;111(3):441–8. Available from: http://link.springer.com/10.1007/s00421-010-1665-6
20. Fox J. A Basic Introduction to Missing Data [Internet]. 2014. Available from: https://socialsciences.mcmaster.ca/jfox/Courses/soc740/Missing-data-notes.pdf
21. Jones A. Age-Grade-Tables [Internet]. 2020. Available from: https://github.com/AlanLyttonJones/Age-Grade-Tables
22. Jones A. Creating Road Running Age-Grade Tables. 2020.
23. Arnab R. Stratified Sampling. In: Survey Sampling Theory and Applications [Internet]. Elsevier; 2017. p. 213–56. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780128118481000078
24. McCullagh P, Nelder J. Generalized Linear Models, Second Edition. Chapman & Hall; 1989.
25. Box GEP, Cox DR. An Analysis of Transformations. J R Stat Soc Ser B [Internet]. 1964 Jul;26(2):211–43. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1964.tb00553.x

26. Vansteelandt S, Avagyan V, Mertens K, Thas O. Principles of Statistical Data Analysis. Ghent, Belgium; 2020.
27. Li P. Box-Cox Transformations: An Overview. 2005.
28. Marquardt DW. Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. Technometrics [Internet]. 1970 Aug;12(3):591. Available from: https://www.jstor.org/stable/1267205?origin=crossref
29. Weisberg S. Applied Linear Regression [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2005. (Wiley Series in Probability and Statistics). Available from: http://doi.wiley.com/10.1002/0471704091
30. Fox J, Monette G. Generalized Collinearity Diagnostics. J Am Stat Assoc [Internet]. 1992 Mar;87(417):178. Available from: https://www.jstor.org/stable/2290467?origin=crossref
31. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. 2nd ed. Springer series in statistics. New York, NY: Springer New York; 2009. 282 p. (Springer Series in Statistics).
32. Hu S. Akaike Information Criterion. 2007.
33. Canty A. Multiple Linear Regression. Hamilton, ON, Canada; 2015.
34. Muñoz CS, Muros JJ, Belmonte ÓL, Zabala M. Anthropometric characteristics, body composition and somatotype of elite male young runners. Int J Environ Res Public Health. 2020;17(2).

8 Appendix

Appendix A: Age Grade Performance

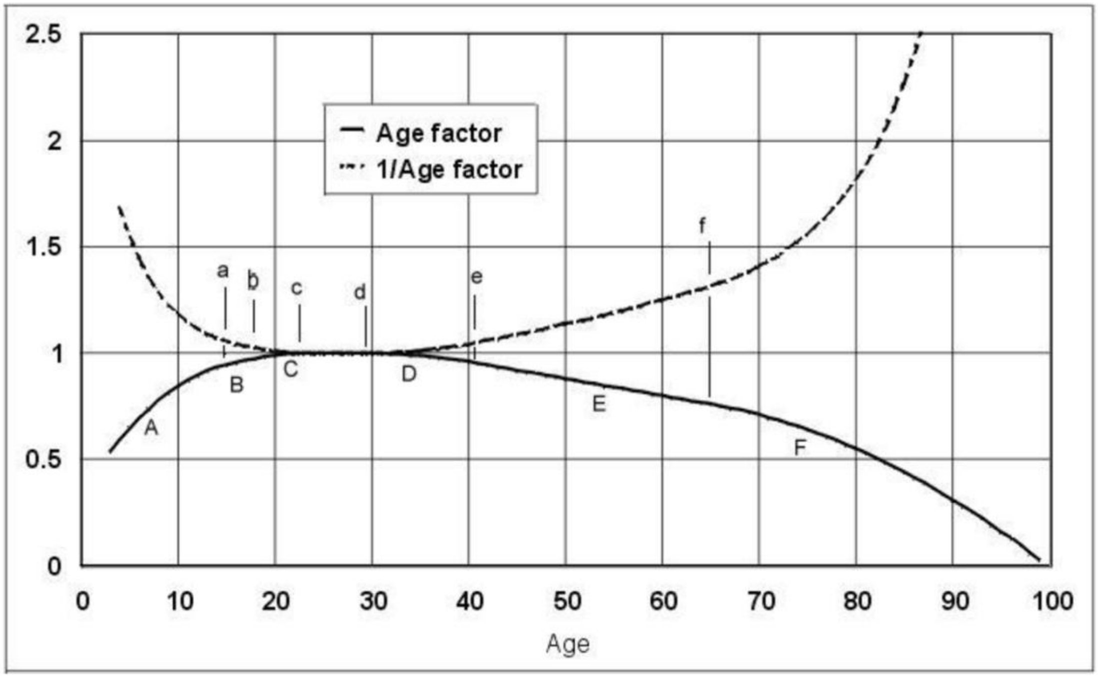


Figure 6: Age grade factors for 5km Event

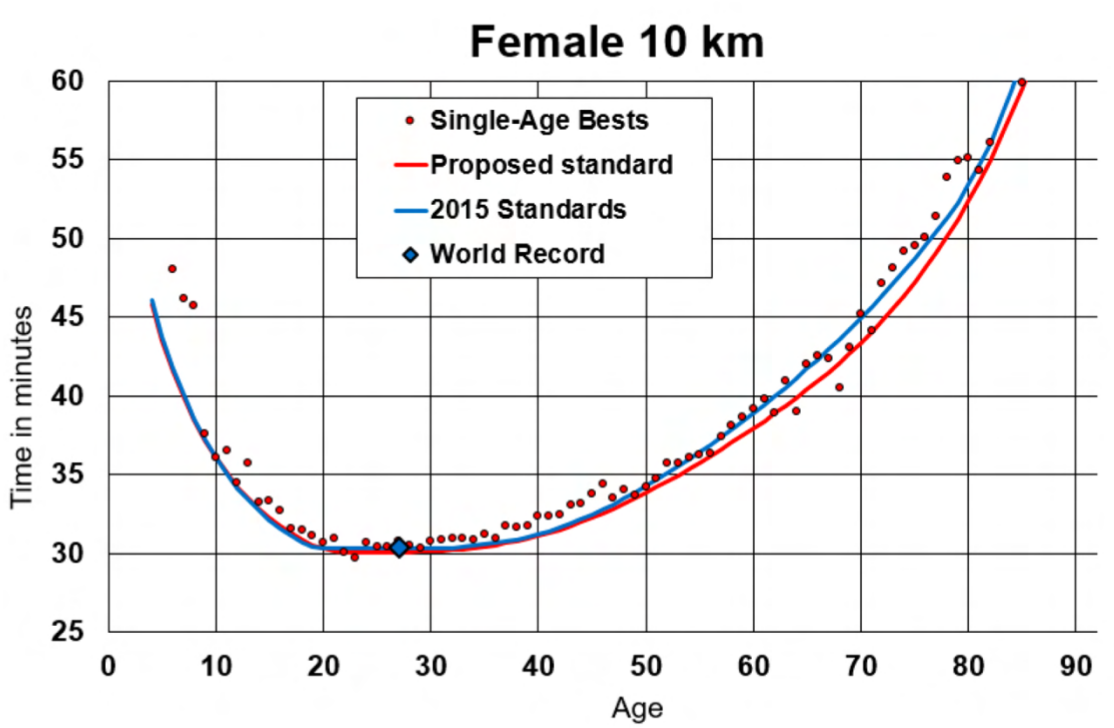


Figure 7: World Records for Female 10km by Age fit with age grade function

Appendix B: Model and Events

Model	Event	N
Running-Hurdle	100m	70
	200m	60
	400m	39
	800m	28
	10 km Road Race (run)	26
	1500m	24
	8 km Cross Country	22
	5000m	18
	80m Hurdles	18
	Half Marathon	14
	10 km Race Walk	10
	20 km Race Walk	9
	300m Hurdles	8
	2000m Steeplechase	7
	5000m Race Walk	7
	110m Hurdles	2
	3000m Steeplechase	2
	400m Hurdles	2
Jumping	Long Jump	33
	High Jump	29
	Triple Jump	13
	Pole Vault	10
Throwing	Javelin Throw	31
	Shot Put	26
	Discus Throw	22
	Hammer Throw	9
	Weight Throw	9

Table 6: Sample size by event and model type

Appendix C: Model Backward Selection Procedures

Running-Hurdle Model

Variables	Full	Model 2	Model 3	Model 3.5
Height	***	***	***	***
Age	***	***	***	***
Gender	***	***	***	***
Weight	***	***	Eliminated	***
Event	***	***	***	***
PmaxTotRelCMJ	***	***	***	***
hminCMJ	**	**	***	*
Ftotmaxg_Hop	**	**	***	**
BFM	*	*	***	Eliminated
PhA		Eliminated	Eliminated	Eliminated

*Table 7: Significance levels for backward selection procedure for Running-hurdle model ***0.001, **0.01, *0.05, '-' 0.01. e=eliminated*

Model	Adjusted-R^2	Standard Error	AIC	5-fold CV MSE
Full	69.29%	0.067	-710.794	0.004
2	69.4%	0.069	-712.6937	0.004
3	69.37%	0.066	-713.3202	0.004
3.5	68.79%	0.067	-707.9944	0.004
Reduced model	64.12%	0.072	-670.7994	0.004

Table 8: Backward selection procedure for Running-hurdle model

Jumping Model

Variables	Full Model	Model 2	Model 3	Model 4	Model 5
Height	***	***	***	***	***
Age	***	***	***	***	***
Gender					eliminated
Weight	***	***	***	***	***
Event	***	***	***	***	***
PmaxTotRelCMJ	***	***	***	***	***
hminCMJ				eliminated	eliminated
Ftotmaxg_Hop	**	**	**	**	**
BFM		eliminated	eliminated	eliminated	eliminated
PhA			eliminated	eliminated	eliminated

*Table 9: Significance levels for backward selection procedure for Jumping model ***0.001, **0.01, *0.05, '-' 0.01. e=eliminated*

Model	Adjusted-R^2	Standard error	AIC	Training MSE
Full	85.69%	0.049	-256.7621	0.002
2	85.72%	0.049	-257.7492	0.002
3	85.74%	0.049	-258.7014	0.002
4	85.8%	0.049	-259.9394	0.002
5	85.77%	0.049	-260.6384	0.002
Reduced model	81.97%	0.055	-241.4023	0.003

Table 10: Backward selection procedure for Jumping model

Throwing Model

Variables	Full	2	3	3.1	4	4.1	5	6	7
Height	***	***	***	***	***	***	***	***	***
Age	***	***	***	***	***	***	***	***	***
Gender						e	e	e	e
Weight			e		e		e	e	e
Event	***	***	***	***	***	***	***	***	***
PmaxTotRelCMJ	***	***	***	***	**	***	***	**	**
hminCMJ	***	***	***	***	-			e	e
Ftotmaxg_Hop	***	***	**	***		*	*		e
BFM		e		e	e	e	e	e	e
PhA		e	e	e	e	e	e	e	e

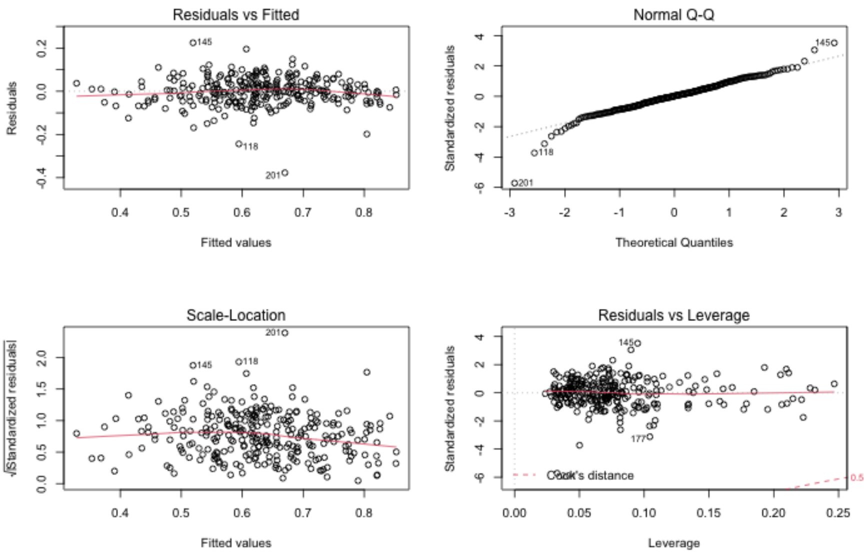
*Table 11: Significance levels for backward selection procedure for Throwing model ***0.001, **0.01, *0.05, '-' 0.01. e=eliminated*

Model	Adjusted – R^2	Standard Error	Training MSE
Full	59.06%	0.064	0.003
Full - log response	60.10%	0.205	0.004
2	59.40%	0.063	0.003
3	59.65%	0.063	0.004
3.1	55.50%	0.066	0.004
3.2 - log-link glm	-	-	0.003
3.3 - log response	60.65%	0.204	0.004
4	44.17%	0.074	0.005
4.1 - log response	49.93%	0.23	0.005
5	41.71%	0.076	0.005
6	42.21%	0.076	0.005
7	42.16%	0.076	0.005
7.1 - log response	45.46%	0.24	0.006
7.2 - log link glm	-	-	0.006
8	42.68%	0.075	0.005
8.1	44.28%	0.074	0.005
9	45.42%	0.074	0.005
Reduced - log response	38.79%	0.254	0.006

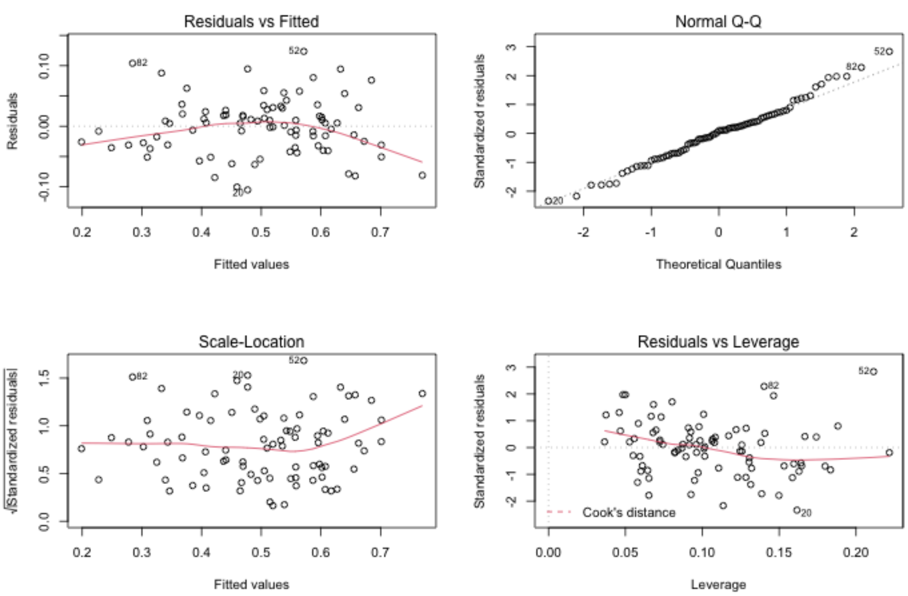
Table 12: Backward selection procedure for Throwing model

Appendix D: Model Diagnostics

Run Diagnostics



Jumping Diagnostics



Throwing Model