BM Musingi
a) Animal breeding and genomics group, Department of Animal Sciences Egerton University, P.O Box 536, 20115 Egerton, Kenya
b) Department of Biological Sciences Egerton University, P.O Box 536, 20115 Egerton, Kenya

N Kitaka
Department of Biological Sciences Egerton University, P.O Box 536, 20115 Egerton, Kenya

G Ogondo
Department of Biological Sciences Egerton University, P.O Box 536, 20115 Egerton, Kenya

TK Muasya
Animal breeding and genomics group, Department of Animal Sciences Egerton University, P.O Box 536, 20115 Egerton, Kenya

LM Mahianyu
National Police Service, Department of Kenya Police P.O Box 49, 20107 Njoro, Kenya

DM Musingi
Department of building and civil Engineering Water section Technical University of Mombasa, Tudor Tom Mboya street P.O Box 90420, 80100 Mombasa, Kenya

Evaluation of modelling in fish length-weight relationships: The current trends

BM Musingi, N Kitaka, G Ogondo, TK Muasya, LM Mahianyu and DM Musingi

Abstract
Length and weight data for three species of fish *Oreochromis niloticus, Lates niloticus*, and *Sydontus victorianus* was collected from three East African lakes, Lake Victoria, Lake Baringo and Lake Naivasha. Correlation were done on weight and length based on the fish species, the lake collected from, sex, and the age. Correlation was found not to be a good measure of evaluating length weight relationship. Models had to be applied. Linear models, random regression models, proportional hazard models and threshold models. Random regression model was found to capture most of the information though it excluded the age of the fish on mean sum of squares error. The challenge with this is that it's a long-term model which accounts for hatching to last date of the fish life.

Keywords: Fish length, fish weight, correlations, models

1. Introduction
Fisheries sub sector in Kenya account for 5.3 billion annually for exports (Ngugi and Manyala 2009) [1]. The supply chain requires big fish in terms of fork length. In Kenya is however not clear the best model to come up with weight of fish of immediate concern is the market segregation where fish are sold according to size of the length. The mature and long big fish are sold in big hotels and export markets. The middle-sized ones are sold to middle class homes and middle-class hotels and restaurants. The small size fish are consumed in slums and the poor marginalized areas (Obwaga, 2020) [2]. There are three main types of fish sold in Kenyan market *Oreochromis niloticus, Lates niloticus* and *Synodontis victorinous*. These fish differ in length and maturity and other stages of growth. Some grow slowly or fast depending on pond-year-season (PYS) factors (Falconer and Mackey 1996) [3]. There is less control of harvesting fish in the inland waters. This means that fish are harvested at any stage be it young, intermediary or mature due to lack of legislative measures from the government (Orina, 2011) [4].

The Kenyan fish population (Nyamweya 2017) [5] reported a decrease in Lake Victoria fish size landings by a coefficient of 0 to about 3.5 from 2008. The proportion of immature fish seems to be rising by 1% per annum in every year landing (Nyamweya 2017) [5]. This calls for study of the effect of the harvesting and a model for length weight relationship that should be easily done for someone to conclude that a fish should not be harvested hence the need to come up with a modelling structure for the best time to harvest. The parameters to be considered in harvesting should be outlined because the main factors that have been used is the length and the weight.

2. Materials and methods
Fish were randomly harvested from L.Naivasha, L.Baringo and L.Victoria. The following were noted from each, number of fish harvested per lake. The species harvested from each lake, the age per species harvested in the lake, the sex of each species harvested in the lake, the length of each species harvested in the lake, the weight of each species harvested in the lake. The data was collected from August 2018 to January 2020. In this period of time, the rainfall was above the mean average of all zones. Basically, all lakes were flooded. The animal breeders tool kit (ABTK) (Golden, 1992) [6] was used to analyze the means and standard deviation and coefficient variations for each of the above traits.

Corresponding Author:
BM Musingi
a) Animal breeding and genomics group, Department of Animal Sciences Egerton University, P.O Box 536, 20115 Egerton, Kenya
b) Department of Biological Sciences Egerton University, P.O Box 536, 20115 Egerton, Kenya

A total of 159 fish was harvested from all the lakes.
Lake Baringo had 56, Lake Naivasha 53 and Lake Victoria 50 species of fish.
There length and weight were immediately measured. Only one type of length was taken which is the total length. Total length (TL) is the length from the most anterior front part to the most posterior point. In this case the tail of the fish is pressed straight to give the longest length. This is as compared to fork length (FL). This is the length from most anterior notch in the length of the tail. In fish like tilapia where the tail has a triangle shape it's the inner most part of the tail in the middle of the tail. For fish without a fork tail like cat fish the FL and TL are the same. The standard length (SL) is the length from anterior point to posterior end of the caudal peduncle. In this study total length was used because is the most common measurement used in fish studies as this is the measurement used in decision setting such as minimum length.

Table 1: Mean weight and mean length and their co-rrelation for lakes

Location	No of Records	Mean Length	Mean Weight	Correlation
Lake Baringo	56	1.915	19.836	1.0
Lake Naivasha	53	2.263	22.450	0.992
Lake Victoria	50	2.237	22.044	0.985

Table 2: Mean length weight relationship per species From the three lakes there were fifty

Location	Location			Total Records	Mean Length	Mean Weight	Correlation
	Lake Baringo	Lake Naivasha	Lake Victoria				
Lates Niloticus	21	15	14	50	19.277	196.853	1.0
Oreochromis Niloticus	19	23	17	59	19.044	212.131	-0.11
Sydontus Victorianus	16	15	19	50	18.947	187.740	0.99
Total	56	53	50	159	57.268	596.724	0.145

Table 3: Analysis of length weight relationship per sex per species

Sex	Sex	Total Number	Mean Length	Mean Weight	Corelation
Lates Niloticus	F	36	18.37	161.39	0.878
	M	15	20.84	273.93	-0.314
Oreochromis Niloticus	F	36	16.43	168.31	1
	M	23	23.52	287.51	-0.222
Sydontus Victorianus	F	27	18.45	157.33	0.852
	M	23	19.69	213.25	-0.83

3. Results and discussion

In Table 1: The mean length for all the lakes was 2.138 with a mean weight of 21.443 at Pearson's Co relation with 95% CI the mean co relation was 0.997. This has also been found in Egypt (Luo 1992) [7] length weight relationship apart from lake Baringo length is not the best indicator for the weight of the fish.

In Table 2: The species showed a different trend. The mean length for the species was 57.268 while the mean weight was 596.724. At 95% CI the correlation was 0.145. This shows species correlation does not play a very big role in estimating weight from length (Panetto, 2010) [8].

In Table 3: There is general agreement that the mean of the female weight is less in all lakes that can be accounted by mouth brooding. The correlation thou between length and weight show that male length is not directly corelated to the weight because female length weight correlation is 0.91 and the other one for males is -0.455. (du Toit, 2012) [9]

From the above data analysis correlation or a straight-line analysis does not account for length weight relationship (Willis 2009) [10]. The relationship between length and weight is not linear (Derek 2013) [11]. This can be explained by the fact that length is a linear measure while weight is related to volume hence as the fish adds linear amount of length is also adding an oval volume connected to its weight. Secondly the availability in weight will increase in shorter fish than in longer fish for example *Sydontus victorianus*. These two above characteristics make the length weight data violate linearity and homoscedasticity which all simple linear regression follow that is y=mx+c. older scientists used to transform this to natural logarithms (Brackwel 2000) [12]. The logarithms will just change to a linear model. This makes most errors additive. This cause for new length weight modelling skills.

3.1 Modelling for length weight relationship in fish

The global trend in fish length weight modelling is now to include all physiological important traits in the model objective. Functional traits such as longevity, disease, fertility, feeding, sex type traits are now featuring prominently in the modelling programs (Zvandilova and Stickova, 2009; Banga *et al.,* 2014; Interbull, 2014) [13, 14, 15]. This is because the economic value of length weight relationship is reported to be up to half of that of production traits especially in aquaculture systems (Jairath and Dekkers, 1994) [16]. In the quest to include many factors in the model, the first step has been to come up with a definition of length weight. To date, length weight relationship is variously defined as length weight of productive life of the fish, lifetime of the weight addition of the fish, the length of aquaculture commodity that is how long fish stays in a pond, total number of broodings and survival from hatching or first brooding to a certain age (Vacek, 2006; Varona *et al.,* 2012) [17, 18]. The second challenge has been the choice of an appropriate analysis procedure. Here the challenge is basically because factors affecting length weight relationship change over time.

Characteristically, length weight is a threshold trait, and displays distinct categorical phenotypes. Threshold traits are influenced by an underlying continuous liability of many genes (Falconer and MacKay, 1996) [3]. For length weight data most fish in aquaculture systems should found in early before brooding or gonadal development, hence the data is

usually skewed to the left (Caraviello, 2004) [19]. Factors that influence length weight relationship especially in a pond situation differ over time depending on prevailing environmental conditions and are therefore time dependent (Zavadilová, 2011; Flynn, 2012) [20, 21]. Survival data can be censored or uncensored. Events like death or culling may be known to have occurred and therefore will be uncensored. On the other hand, the events such as culling or death may not have been recorded because they are not known to have occurred because fish are small animals and might not be easily culled based on size. If fish are alive at the time of analysis, they will be included in the evaluation and are therefore regarded as censored (Beswick, 2004) [22]. These complex characteristics or features of survival data require careful consideration and choice of the appropriate modelling strategies in order to discover all phenotypic, additive and environmental variance (Weigel, 2003) [23]. Approaches used in the analysis of length weight relationship include linear models (Du Toit, 2011; Kern, 2014) [24, 25]. This method has been used for length weight relationship measures defined quantitatively such as fish weight, age from hatching or brooding to last day of harvest and total days in brooding or milting. Other models include random regression (Van Pelt and Veerkamp, 2014) [26, 33].

3.2 Linear models
This approach accounts for censored data or records of fish which are still alive at the time of analysis. Survival of a fish to a predefined age or time period is used. The time period or age t, is pre-determined on a time scale and the record of each fish is assigned a 1 or 0, for successful survival to that age or not, respectively. The model is as follows:

$$y_{ij}(t) = \mu + s_i + e_{ij}$$

where $y_{ij}(t)$ =0 if the jth progeny of the ith sire (male fish) did not survive to time *t* and 1 if it was alive at that time. μ is the overall popualtion mean, s_i is the breeding value of the ith sire on the binary scale and e_{ij} are random residuals. This approach to modelling length weight relationship has been applied to Czech Fleckvieh (Zavadilová, 2009b) [13], South African catfish (Du Toit, 2012) [35] among other fish populations in aquatic systems (Samore, 2010; Zavadilová and Štípková, 2012) [27, 36]. This approach is simple and can accommodate univariate as well as multivariate length weight relationship, sire or maternal grandsire models. The drawback of this model is that it makes unrealistic assumption that data are continuous and are normally distributed which is not common in fish because of high mortality rate and other effects like floods (Yazdi, 2002) [28]. Fish that are at different ages before the pre-determined time point are treated the same way, leading to erroneous results (Ducrocq, 1988) [29]. Factors that influence length weight relationship over the time period are not considered such as disease, since survival times are derived from a product of all measurements of either location, sex, maturity level or species of the fish (Vukasinovic, 1999) [30].

Linear models are more appropriate in analyzing continuous traits such as weight addition rather than binary traits. This is the caution that was taken by (Kern 2014) [25]. in this study measure of length weight relationship which were continuous e.g. lifetime brooding, age from hatching or first brooding to last day in brooding, total days in brooding over all broodings the fish did were analyzed using linear models, while measures of length weight relationship related to survival were analyzed using threshold models. This was necessary because use of linear model BLUP (best linear unbiased prediction) was inappropriate because with categorical, data like length and weight values and residuals would not be independent of each other (Gianola, 1982) [31].

3.3 Random regression models
Random regression models model additive genetic values as a function of an observed dependent variable through a set of random coefficients. In modelling survival data using random regressions, fish records are assigned binary units (0 or 1) if it survived a brooding or a month after brooding. A linear model with random regressions for additive fish effect as random effects is then fitted for genetic evaluation and length weight relationship values generated for any point in the trajectory (Jamrozik., 2008; Van Pelt, 2015) [26, 32]. A univariate random regression model is described as follows:

$$y_{ijklmno:t} = (YS:Ht)_{ij} + (YSAP:t)_{ikl} + r(a, xm1)_n + r(pe, x, m2)_k + e_{ijklmno:t}$$

where $y_{ijklmno:t}$ is the nth observation of the kth fish at time t of the ith fixed effect and jth group, YS is the ith year-season of first brooding, H is the jth pond, A is the kth age at first brooding class, P is the weight level, $y_{ijklmno:t} \sum_{i=0}^{m_1} a_{kl} x_{ijk:l}$ is notation for random regression, where a is the additive genetic effect for the kth fish, x are orthogonal polynomials of time t, after brooding, *a* is the random regression coefficients for additive genetic effects for fish n labelled using a bead, *pe* are the random regression coefficients for permanent environmental effects for fish n example pond temperature and maternal genetic effects and $e_{ijklmno:t}$ are random residuals. Random regression models are more robust to effects of censoring data (Veerkamp, 1999) [33] and are closer to proportional hazard models and generalized linear models. They can also handle multiple traits (Jairath, 1994) [16]. This is what is recommended for the above data because it accounts for location for aquatic systems, species of the fish, sex of the fish, maturity of the fish and the environmental factors.

3.4 Proportional hazard models
Proportional hazard (PH) models model survival as the probability that fish will survive past a specified time *t*, and the hazard function, which is the instantaneous rate of failure or death (Ducrocq, 2005) [37]. The survival function and the hazard functions in PH modelling differentiates between a fish that dies exactly at time *t*, and one that was last alive at time *t*, and may have survived on. PH models make the assumption that the hazard rate or risk is a function of time-dependent baseline hazard and an exponential function of a series of explanatory variables i.e. covariates example fish from the same parents which is a common trend in fish (Ducrocq, 1997) [38]. PH models may not result in the best fit to data and therefore may not accurately estimate the performance of future fingerlings of an individual. They also make the assumption that survival is the same trait throughout the lifetime, although correlations between broodings 1 to 3 have been reported to be less than unity. Another limitation is

their inability to account for non-random mating among fish and cannot handle multiple traits. Despite these limitations they are able to handle censored data and can accommodate non-normal data distribution well and incorporate time-dependent environmental effects like floods and diseases (Ducrocq, 1988) [29]. Common PH models include Weibull and Cox models, with the former being more popular.

3.5 Threshold models
In threshold modelling, survival is considered as a binary trait (0=dead at time *t* and 1=alive at time *t*). Threshold models include sequential threshold models, threshold repeatability models, and threshold cross-sectional models. Threshold models have been used to evaluate survival to first hatching in tilapia survival in catfish. Threshold models have been used to analyze measures of length weight relationship related to survival such as survival from hatching or first brooding to predetermined age (Kern, 2014) [25]. The major limitation of threshold models is that heritability estimates are from an underlying continuous scale, and their effect on rate of genetic gain may be similar to that from linear models (Boettcher., 1999) [34]. However, they are capable of handling multiple traits and can handle large datasets of over 10,000 fish.

3.6 Model comparison
Linear, threshold and random regression models yield lower estimates of length weight relationship, ranging from 0.01 to 0.18 (Kern, 2014; van Pelt and VeerKamp, 2014) [25, 26, 39] than proportional hazard models (0.15 to 0.22) on the original scale. However, the former group of models is able to accommodate multiple traits and therefore can produce genetic correlations of length weight relationship with other indicator traits like age of the fish and sex (Kern, 2017) [25]. found that threshold models yielded higher length weight relationship estimates (0.04) compared to those obtained from linear repeatability and cross-sectional models (0.02). In terms of accuracy of evaluations, the models were compared based on the correlation between true and estimated length weight values for proportion of sire's daughters that survived to a specified age and sire ranking. The correlation for Weibull models remained constant over time, but increased over time for linear and random regression models, though they were lower than for the former models. The models also differed in terms of ranking of sires. For instance, linear models and Weibull models had similar sire rankings compared to proportional hazard models. Multi-trait models were also found to be superior in terms of predicting survival of female fish up to a certain age compared to Weibull models with Weibull models showing poor prediction of proportion of daughters of sires in early fish lifetime than multiple trait models.

4. Conclusion
In general, linear and threshold models have been found to be superior to Weibull models when fish length weight values were estimated when predicting survival to 150 days from first brooding Weibull models are better when predicting functional length weight relationship while linear random regression and linear multiple trait models are the models of choice when predicting overall survival. Linear models are however favored when studying the predictive ability of sire estimated breeding values of survival to a particular age, since this is better measured by average length weight relationship instead of functional length weight relationship. Threshold

models are more appropriate for analyzing binary survival than linear models while linear multiple trait and random regression models resulted in inferior statistical correctness (inferior model fit). Threshold length weight relationship models also result in higher length weight estimates than linear male models.

5. References
1. Ngugi, Manyala. Analysis of fish growth in Kenya ARS Washington DC. 2009. (6)118-119.
2. Obwaga B, Soma K, Rurangwa E, Kilelu C. Resilient robust and reliable agro-food sectors in Kenya Wageningen UR, 2020.
3. Falconer, MacKay. Introduction to quantitative genetics fourth edition Longman Scientific and technical, Harlow UK, 1996.
4. Orina B, Munguti J. An overview of Kenyan Aquaculture. Fish Aquat Sci, 2011; 19(1).
5. Nyamweya K, Manyala D. Rise of aquaculture in Kenya. International Journal of Aquatic and Fisheries Studies. 2017; 1(6).
6. Golden Animal breeders tool kit user's guide and reference manual, 1992.
7. Luo M, Yazdi R, Assessment of Egyptian fish; Journal of East African studies. 1992; (6):10-12.
8. Panetto R, Banga C. Assessment of fish breeding systems. journal of fish science. 2010: (93):112-124.
9. Du Toit Genetic evaluation of aquatic productive life. Phd dissertation University of Ghana, 2012.
10. Willis R, Delphos K. Reliability of genetic evaluations in fish. Journal of fish science. 2000; (96):40-49.
11. Derek M, Peters MA. test of quantitative genetic theory of tilapia. Journal of evolutionary biology. 2013: 18;840-845.
12. Brackwel R, Peters K. In depth pedigree analysis of tilapia. Genetics and molecular research. 2014: (4):125-130.
13. Zavadilová, Stipakova. Bayesian measures of model complexity and fit Journal of America stastitics. 2009; 65:593-624.
14. Banga C, Khumelo D. Vertm South African fish Association, 2014.
15. Interbull. Analysis of phenotypic relationship in fish. Journal of Czech fish science, 2014.
16. Jairath, Dekkers. Genetic evaluation of aquatic systems in Europe. Journal of fish science. 1994; 10:650-653.
17. Vacek Pacek. Genetics of the eel Journal of East Africa genetics. 2006; (1):10-17.
18. Varona, Stipacova. Relation between length and weight in fish. Czech journal of fish science. 2012; 91:537-543.
19. Caraviello C, Paceto M. Analysis of relationship between length and weight in eel fish. Genetic selection and evolution. 2004; 88:91-96.
20. Zavadilová, Stipacova. Models and implications of fish science Journal of International statistics. 2011; (7):118-121.
21. Flynn, Effect on genetic correlations in fish. Journal of fish science. 2012; 98:1256-1265.
22. Flynn Quantitative analysis of fish genetics. Journal of European fisheries. 2012; (6):220-227
23. Beswick, Lambert. Review: statistics review.12: survival analysis critical care. 2004; 8:389-394.
24. Weigel, Wescick. Investigation of factors affecting over cultural culling in Australia, Australian journal of fish

science. 2003; 86:905-915.

25. Du Toit, Khumelo. Genetic evaluation of fish ponds in South Africa for performance and functional traits South African journal of fish science. 2011: 45:103-109.

26. Kern Kern. Proportional Hazard Modelling In Fish Scientia Agricola. 2014, 2014b:45:85-90.

27. Van Pelt, Veerkamp. Genetic evaluation of pond life. Inter fish bulletin, 2014.

28. Samore Zavadilová, Štípková. Genetic parameters for functional traits in fish. Italian journal of fish science, 2010, 2012, 20, 546-564, 22, 220-227.

29. Yazdi, Luo. Evaluation and response to selection in fish using proportional hazard models. European journal of fish science. 2002; 158:202-209.

30. Ducrocq. Mixed model analysis for length and weight in fish 10th world congress on fish genetics. 1988; 42:510-515.

31. Vukasinovic, Veerkermp. Comparing models for fish science inter fish bulletin. 1999: (21):3-10.

32. Gianola, Burton. Brooding and its effect on length weight relationship in tilapia. Journal of fish science. 1982: (61):439-448.

33. Jamlozick, Bugger. Kurmer Scientific evidence of fish weight European Journal of fish science. 2012; 18:312-319.

34. Veerkemp, Vanpelt. Modelling skills in fish farming Scotland Journal of fish research. 1999; (7):81-89.

35. Boettcher, Burton. Trends of fish harvesting Asian Journal of fish studies. 1993; (8):101-110.

36. Du Toit. Assessment of fish length weight relationship in ponds in South African Journal of fish science. 2012; 42:65-67.

37. Zavadilova, Stipakova. Modelling skills in animal research Scotland Journal of research. 2012; (10):101-119.

38. Ducroq. Advanced modelling techniques Canadian science. 2005; (11):101-115.

39. Ducrocq. Biological modelling American Journal of science. 1997; (15):91-106.

40. Veerkamp, Vanpelt. South African Journal of genetics. 2014; (8):91-104.