# EVALUATING ALTERNATE MODELS TO ESTIMATE GENETIC PARAMETERS OF ECONOMIC TRAITS IN *CLARIAS MAGUR*

Thesis submitted in partial fulfillment
of the requirements
for the degree of

**Ph.D. (Fish Genetics)**

by

**RAMEEZ ROSHAN, P.M., M.F.Sc.**
**(FGB-PA5-03)**

**ICAR-CENTRAL INSTITUTE OF FISHERIES EDUCATION**
(Deemed-to be University)
(Established Under Section 3 of UGC Act 1956)
**Panch Marg, Off Yari Road, Versova**
**Andheri (W), Mumbai – 400061**

**FEBRUARY 2021**

**Rameez Roshan, P.M., 2021**. Evaluating alternate models to estimate genetic parameters of economic traits in *Clarias magur*. Ph.D. Thesis, ICAR – Central Institute of Fisheries Education (University Established Under Section 3 of UGC Act 1956) Panch Marg, Off Yari Road, Versova, Andheri (W), Mumbai – 400061.

**ISBN: 978-81-951963-4-0**

# *Dedication*

*My Mother Sajitha M Kunju*

*and*

*My Father Muhammed Kunju P.M.*

*For their endless love, support and encouragement*

भा.कृ.अनु.प.- केन्द्रीय मात्स्यिकी शिक्षा संस्थान
ICAR-CENTRAL INSTITUTE OF FISHERIES EDUCATION
(A University Established Under Sec. 3 of UGC Act 1956)
Ministry of Agriculture & Farmers Welfare,
Govt. of India.

Dated: 26 February, 2021

# CERTIFICATE

Certified that the thesis entitled "**EVALUATING ALTERNATE MODELS TO ESTIMATE GENETIC PARAMETERS OF ECONOMIC TRAITS IN *CLARIAS MAGUR***" is a bonafide record of independent research work carried out by **Mr. Rameez Roshan, P.M.** during the period of study from September, 2015 to February 2021 under our supervision and guidance for the degree of **Doctor of Philosophy (Fish Genetics)** and that the thesis has not previously formed the basis for the award of any degree, diploma, associateship, fellowship or any other similar title.

**Chair**

**(Shrinivas Jahageerdar)**
Principal Scientist
Fish Genetics and Biotechnology Division
ICAR-CIFE, Mumbai, India

**Advisory Committee**

**(J. Jayasankar)**
Principal Scientist
FRAD Division
ICAR-CMFRI, Kochi, India

**(Rama Bangera)**
Senior Geneticist
Benchmark Genetics
Bergen, Norway

**(Thongam I. Chanu)**
Scientist
Aquaculture Division
ICAR-CIFE, Mumbai, India

पंच मार्ग, ऑफ यारी रोड, वरसोवा, अंधेरी (प), मुंबई - ४०० ०६१. (भारत)
Panch Marg, Off Yari Road, Versova, Andheri (W), Mumbai - 400 061. (India)
कार्यालय / Office : 022-26361446/7/8,
Fax : 022-26361573
Website : www.cife.edu.in

# DECLARATION

I hereby declare that the thesis entitled **"EVALUATING ALTERNATE MODELS TO ESTIMATE GENETIC PARAMETERS OF ECONOMIC TRAITS IN *CLARIAS MAGUR*"** is an authentic record of the work done by me and that no part thereof has been presented for the award of any degree, diploma, associateship, fellowship or any other similar title.

Date: 26 February, 2021
Place: Mumbai

(Rameez Roshan P.M.)
Ph.D. Scholar
ICAR-CIFE

# *Acknowledgments*

*Alhamdulillah, for all the concealed and visible guidance at every stage of my Ph.D. research, I express my sheer incapacity of thanking The Almighty.*

*I want to take this unique opportunity to express my most profound sense of gratitude and indebtedness to my esteemed guide Dr. Shrinivas Jahageerdar, Principal Scientist, Fish Genetics and Biotechnology Division, ICAR-CIFE, Mumbai, for suggesting this research topic and for his indefatigable effort, constructive criticism, erudite counseling, instinctive attention, constant encouragement, inspiring suggestions, benign attitude and sagacity coupled with patience in guiding and caring for me at every stage of this investigation and preparation of this manuscript. I also thank him from the core of my heart for giving me time from his busy schedule.*

*My sincere and heartfelt thanks are to my advisory committee members Dr. J. Jayasankar, Principal Scientist, ICAR-CMFRI, Kochi, India, Dr. Rama Bangera, Senior Geneticist, Benchmark Genetics, Bergen, Norway, and Dr. Thongam Ibemcha Chanu, Scientist, ICAR-CIFE, Mumbai, India, for their continuous support and worthy suggestions.*

*I am incredibly thankful to Dr. Gopal Krishna, Director, ICAR-CIFE, Mumbai, for his support in all forms. I thank Dr. Aparna Chaudhari, Principal Scientist and Head (Acting), FGB Division, and all the faculty and technical staff of FGB Division for all the help and support. I would also like to record my heartfelt thanks to Dr. Annam Pavan Kumar and Dr. Mujahid Khan Pathan, Scientists, FGB Division, for their moral support and constant encouragement. I also wish to thank Mr. Snajeev, Mr. Sanjay and Mr. Jayprakash Patil for their continuous help.*

*I want to place on record my special gratitude and thanks to Dr. Arthur Gilmour, Statistical and ASREML consultant, Australia, for his valuable inputs. I also want to express my thanks to Dr. A. Dhandapani, Principal Scientist, ICAR-NAARM, Dr. Arnold Saxton, Professor Emeritus, University of Tennessee, USA, Dr. Karin Meyer, Professor, University of New England,*

*Australia and* **Dr. Bruce Walsh**, *Professor, University of Arizona, the USA for helping me sort out issues with their expertise at the time of mid-air hiccups.*

*I want to express my sincere thanks to* **Dr. Arun Sharma**, *Scientist, ICAR-CIFE, Mumbai,* **Shri Srinivas Rao, Shri J. Krishna Prasad** *and* **Mr. Sambamurthy Garu** *for their untiring efforts and help during data collection and all the Technical staff of ICAR-CIFE, Balabhadrapuram Centre, for their immense support during the field experiment. I also wish to acknowledge* **Dr. Muralidhar P. Ande**, *Scientist, and OIC Kakinada Centre for the help.*

*I want to place on record my special gratitude to* **Dr. Syed Ismail Bukhari Thanggal, Maulana Najeeb Usthad, K.M Sidhic Maulavi** *for being the backbone of my family, especially when I was away with the Ph.D. research.*

*I am at a loss of words to express my sincere gratitude and indebtedness to my dearest and closest friends* **Dr. Sayeda Neha Wajahat Qureshi** *and* **Dr. Shamna N.,** *for their moral support, constant encouragement, valuable suggestions, and for taking care of me during my hard times.*

*I acknowledge the support of my beloved juniors* **Mr. Thushar P. Kumar, Mr. Dayal Devdas**, **Mr. Shyam Kaaduvetti, Mr. Suman Dey, Ms. Saumya Pandey, Ms. Akshaya,** *and* **Ms. Nuzaiba P.M.**

*I consider it my proud privilege and immense pleasure to acknowledge my loving family, who has been my unshakable support during the entire course of my life, including this piece of research work, though words are not sufficient and rather not necessary to express my gratitude. I wish to express my heartfelt thanks to my beloved sister* **Mrs. Risana Parveen P.M.** *and my Brother in Law,* **Mr. Sahal Usman,** *for their support.*

*Financial Assistance provided by the* **Indian Council of Agricultural Research** *through fellowship is greatly acknowledged. I also would like to acknowledge the* **National Agricultural Higher Education Project**, *ICAR-CIFE, to award me the Research Associateship.*

*Date: February 26, 2021*
*Place: Mumbai*

*(Rameez Roshan, P.M.)*
*ICAR - CIFE*

# सारांश

सभी प्रकार की जलकृषि संभावनाओं सहित क्लैरियस मागूर यह भारतीय कैटफिश प्रजातिओं में से एक प्रजाति है । वर्तमान में, मागुर जलकृषि, बीज की उप-अनुकूलतम गुणवत्ता एवं मात्रा द्वारा नियंत्रित है । मागुर की जलकृषि मुख्यत: प्राकृतिक प्रजनन आधार से एकत्रित किए गए बीज पर निर्भर है, इसके नि:शेषण के लिए मुख्यरूप प्राकृतिक संचय नमूना पर दबाव दिया जा रहा है । अनुकूलतम मात्रा में गुणवत्ता वाले बीजों की धारणीय आपूर्ति के लिए आनुवंशिक चयन की संभावना है । हालांकि, चयन प्रतिक्रिया के उच्चतम सीमा के लिए आनुवंशिक चयन आशावादी है, विश्वसनीय आनुवंशिक पैरामीटर के अनुमानों की जानकारी होना अनिवार्य है । इसलिए, वर्तमान कार्य का उद्देश्य वैकल्पिक नमूनों और प्रणाली का पालन करने से मागुर में आर्थिक लक्षणों के आनुवंशिक मापदंडों का अध्ययन करना है । गैर-आनुवंशिक कारक बैच, स्टॉक, तालाब और सेक्स का मागुर की वृद्धि पर महत्वपूर्ण प्रभाव पड़ा । यह संवर्धन प्रकार पैदावार पर शरीर के वजन के लिए सार्थक सहायक नहीं है, जिसका विकास एक और बहुसंवर्धन दोनों प्रणलियों के लिए बहुविध वंश-कुल पर प्रतिबंधित है । अकेले संचय (स्टॉर्किंग) में शरीर के वजन में विषमता ने पैदावार के शरीर के वजन में 12.5% की भिन्नता का देखी गई, जो कि शरीर के वजन को कम करने की एकरूपता को प्रभावित करने वाले विभिन्न कारकों के प्रबंधन के लिए महत्व का अहसास कराता है । संचय में वृद्धि के लक्षणों की वंशागतता बहुत अधिक थी (0.97-0.74), जो पैदावार की ओर घटती है (0.07-0.44) । वृद्धि लक्षणों के बीच आनुवंशिक सहसंबंध सकारात्मक और उच्च (0.80 से 0.99) थे तथा इसलिए समलक्षणीय (फेनोटाइपिक) सहसंबंध (0.34 से 76) थे । पैदावार की वंशागतता का अनुमान लगाने के लिए BW, ANOVA, REML, पैरामीट्रिक बूटस्ट्रैप, एसिम्प्टोटिक सैंपलिंग, जैकनाइफ, और बायेसियन पोस्टीरिओर का समरुप परिणाम (0.43-0.45) की वैकल्पिक प्रणाली उपयोग की गई । जब कि गैर-पैरामीट्रिक BLUP आधारित बूटस्ट्रैप वंशगतता उच्च थी (0.51) । एकचर और बहुचर नमूनों द्वारा अनुमानित प्रजनन दर के बीच सहसंबंध दर्जा उच्च था (0.80 से 0.99) । अलग-अलग उम्र अवस्था में दर्ज किए गए BW के बहु-लक्षण मूल्यांकन के परिणामस्वरूप नौ महीने और पैदावार के समय BW के बीच एक इकाई आनुवंशिक सहसंबंध हुआ, जिससे कम उम्र अवस्था के चयन की संभावना का पता चला । विसंगति की शेष विविधता के साथ नमूना निकासी अनियमित थी मूल्यांकन सहसंयोजक कार्यों का अनुमान लगाने के लिए द्विघात लीजेंड बहुपद का उपयोग करते हुए विकास प्रक्षेपवक्र विकास के आनुवंशिक मूल्यांकन के लिए सबसे अच्छा आवेश देखा गया । सहसंयोजक कार्यों का उपयोग करते हुए, पचास अलग-अलग उम्र अवस्था के लिए आनुवांशिक मापदंडों का अनुमान लगाया गया था, जिन पर BW रिकॉर्ड उपलब्ध थे, जिसके परिणामों ने आनुवंशिक चयन के माध्यम से औसत विकास प्रक्षेपवक्र को बदलकर मागुर में वृद्धि को बढ़ाने की बहुत अधिक संभावना का सुझाव दिया था । वर्तमान अध्ययन में मागुर की आधारभूत संख्या (पैदावार) में योजक आनुवंशिक विसंगति की अधिक उच्च उपस्थिति का पता चला, जो आनुवंशिक चयन के माध्यम से मागुर के तेजी से बढ़ते उच्च प्रदर्शन तनाव सीमा को प्रमाणित करता है ।

# Abstract

*Clarias magur* is an Indian catfish species with the potential for aquaculture. Presently, the magur aquaculture is constrained by suboptimal quality and quantity of the seed. The genetic selection has the potential for the sustainable supply of quality seeds in optimum quantities. To maximize the selection response, the knowledge of reliable genetic parameter estimates is essential. Hence, the present work aimed to study the genetic parameters of economic traits in magur by implementing different models and methods. The non-genetic factors viz., batch, stock, pond and sex had a significant effect on the growth of magur. The culture type did not contribute significantly to body weight at harvest, which precluded the development of multiple strains for both mono and polyculture systems. The high heritability of body weight (BW) at stocking (0.74) decreased towards harvest (0.44). The genetic correlations between growth traits were positive and high (0.80 to 0.99) and so were the phenotypic correlations (0.34 to 76). Among the different methods used to estimate the heritability of harvest BW, ANOVA, REML, parametric bootstrap, asymptotic sampling, jackknife, and Bayesian posterior gave similar results (0.43-0.45), and non-parametric BLUP based bootstrap heritability was high (0.51). The rank correlations between the breeding values estimated by univariate and multivariate models were high (0.80 to 0.99). The high genetic correlation between BW at nine months and harvest by multi-trait model suggests a possibility of early age selection. A random regression model with heterogeneous residual variance gave the best fit for genetic evaluation of growth trajectories using a quadratic Legendre polynomial to estimate covariance functions. The results of genetic parameter estimate at fifty different ages suggested the high possibility to enhance growth in magur by applying selection on growth trajectories. The presence of high additive genetic variance in the base population of magur, suggests a possibility of developing a fast-growing high-performance strain of magur through genetic selection.

# CONTENTS

# ABBREVIATIONS

| S.No. | Abbreviation | Full Form |
|---|---|---|
| 1 | ADG | Average Daily Gain |
| 2 | AIREML | Average Information Residual Maximum Likelihood Algorithm |
| 3 | ANOVA | Analysis of Variance |
| 4 | AP stock | Andhra Pradesh stock |
| 5 | BD | Body Depth |
| 6 | BLUE | Best Linear Unbiased Estimator |
| 7 | BLUP | Best Linear Unbiased Predictor |
| 8 | BQUE | Best Quadratic Unbiased Estimators |
| 9 | BUE | Best Unbiased Estimators |
| 10 | BV | Breeding Value |
| 11 | BW | Body weight |
| 12 | CF | Covariance Functions |
| 13 | DFREML | Derivative Free Residual Maximum Likelihood Estimation |
| 14 | EM | Expectation Maximization Algorithm |
| 15 | FWFF | Fresh Water Fish Farm |
| 16 | HW | Head Width |
| 17 | K | Condition factor |
| 18 | LMM | Linear Mixed Model |
| 19 | LP | Legendre polynomials |
| 20 | MCMC | Markov Chain Monte Carlo |
| 21 | ML | Maximum Likelihood Estimator |

| S.No. | Abbreviation | Full Form |
|---|---|---|
| 22 | MME | Mixed Model Equations |
| 23 | MSS | Mean Sum of Squares |
| 24 | MVN | Multivariate Normality |
| 25 | Non AP stock | Non Andhra Pradesh stock |
| 26 | NR | Newton Raphson Algorithm |
| 27 | PEV | Prediction Error Variance |
| 28 | PIT | Passive Integrated Transponder |
| 29 | REML | Residual Maximum Likelihood Estimator |
| 30 | RRM | Random Regression Models |
| 31 | SE | Standard Error |
| 32 | TL | Total Length |

# LIST OF TABLES

# LIST OF FIGURES

# 1. INTRODUCTION

*Clarias magur* (Hamilton, 1822) is an Indian catfish species popularly known as magur, is widely distributed across India, Nepal, Bhutan and Bangladesh among other countries (Ng and Kottelat, 2008). Magur has the potential to be a candidate species for freshwater aquaculture in India. The lack of good quality seed in required quantity is the primary constraint to take up the commercial aquaculture of magur. An optimally designed genetic selection program of magur has the potential to supply the desired number of high-quality broodfish to establish commercial hatcheries, which is an important long-term and sustainable way to increase the magur aquaculture in India.

Selective breeding offers an opportunity for continuous genetic gain for traits under selection. Selective breeding programs in aquaculture species have been usually aimed at enhancing growth performance viz., body weight at certain age (usually at the harvest) and reducing the culture period , which leads to more efficient production and higher benefits (Gjedrem and Baranski, 2009). In order to standardize the aquaculture practice and to optimize any breeding programme to obtain maximum response to selection, the knowledge of non-genetic factors affecting the growth traits and their genetic parameters are essential prerequisites (Gjedrem and Robinson, 2014). In magur, limited information is available on non-genetic factors affecting growth traits and their genetic parameters. Jousy *et al*. (2018) estimated the effect of age class, culture type, pond and sex on the harvest body weight of magur. The extent of genetic improvement in successive generations is proportional to the existing genetic variation within the population. Heritability estimates for body weight at harvest and other economically important traits in several aquaculture species have been well documented (Gjedrem, 1983; Gjedrem and Robinson, 2014). In magur, Jousy *et al.* (2018) reported a high heritability (0.63±0.10) for harvest body weight.

The linear mixed model (LMM) analysis, is the popular tool for analysing breeding data where first the variances are estimated followed by prediction of breeding

values. Among different methods of estimation of variance components (Searle *et al*., 2009), Residual/Restricted Maximum Likelihood (REML) is the method of choice in pedigreed selection experiments (Lynch and Walsh, 1998). However, the properties of likelihood estimators are valid only when the sample size approaches infinity (asymptotically), and the behaviour of the same under small sample size is mostly unknown (Psutka and Psutka, 2019). Further, it is difficult to calculate reliable confidence intervals around functions (heritability) of these parameters (Waldmann and Ericsson, 2006). The standard procedure to obtain the sampling variance of heritability is to linearly approximate the function with its first-order Taylor series expansion and then to estimate the variance of this linear approximation by the Delta method (Lynch and Walsh, 1998; Meyer and Houle, 2013). However, the sampling variance might be biased under an incorrect asymptotic approximation, especially when the sample size is small (Thai *et al*., 2013). Under large sample approximations, it is unclear as to what amount of data constitute a large enough sample size (Walsh and Lynch, 2018). No direct approaches are available to estimate the optimum sample size which could give a meaningful asymptotic approximation of likelihood estimates; hence, how large the sample size is often a grey area. Therefore, the use of alternate approaches for the evaluation of uncertainties in variance components and its functions is desirable. Sampling-based methods like non-parametric and parametric bootstrap estimation, jackknife estimation, asymptotic sampling and MCMC based methods are a few options for estimating uncertainties associated with the parameter estimates, which can provide a complete picture of the uncertainties associated with estimating unknown parameters.

Usually multivariate BLUP models predict breeding values with higher accuracies in comparison to the corresponding univariate models (Mrode, 2014) and are preferred when different traits are measured on each animal. However, the traits measured sequentially over time on each animal will constitute repeated measurements which are special form of multivariate case (Van der Werf and Schaeffer, 1997). A multi-trait model can be used for the genetic evaluation of repeated measurements, but, for many different ages a multi-trait model could be less parsimonious due to

overparametrization. Also, the covariance patterns of repeated records are highly structured in such a way that the covariance between them decreases sequentially as the time interval between measurements increases. The unstructured covariance matrix in a multi-trait model cannot account for the change in covariance structure between repeated records as a function of time. Therefore, modelling the trajectories of repeated measurements as a function of parameters, that define those trajectories, can give information on the effects that causes variation among the trajectories. Kirkpatrick and Heckman (1989) defined covariance functions as a means to account for continues change of covariance with time, which utilize the information about parameters that define trajectory. The growth trajectories are economically important trait and the parameters that describes typical features like intercept, slope and curvature of the random variation in growth trajectories can be estimated using random regression models. The covariance between the random regression coefficients can be used within covariance functions to estimate genetic parameters between any time points within a given interval (Schaeffer, 2016). Hence, multi-trait models and random regression models can be used as an alternate methods to estimate genetic parameters.

The goal of this study was to evaluate the genetic potential of *Clarias magur* by employing different models and methods to estimate reliable genetic parameters which is essential to optimize the breeding program for maximization of selection response. The specific objectives of the study were

- To quantify the effect of non-genetic factors on various economic traits
- To evaluate the genetic parameters by different models
- To predict the breeding value based on the univariate and multivariate models

# 2. REVIEW OF LITERATURE

*Clarias magur* (Hamilton, 1822), popularly known as magur, belongs to the order Siluriformes and family Clariidae. Magur is distributed across India, Bangladesh, Bhutan, Nepal and other Asian countries (Ng and Kottelat, 2008). Due to the critical decline in the wild population, magur has been classified as an endangered species by the IUCN Red List (2021) (Vishwanath, 2010). The species is considered to be the most promising candidate species for freshwater aquaculture diversification in India (Sahoo *et al*., 2016). The magur aquaculture is practiced across different part of the country, and differential growth patterns are reported. Borah (2020) reported the average length (30 cm) and weight (200 g) of adult wild-caught magur. Sahoo *et al*. (2019) reported the mean length and weight of hatchlings (0.5 cm and 0.0022 g), fry (2 cm and 0.09 g), fingerlings (7.27 cm and 4.13 g), juveniles 11 cm and 18.18 g), and adult (25.26 cm and 138.87 g) from hatchery bred magur. Jousy *et al*. (2018) reported an average harvest body weight of 143.71 g after one year of pond culture. In another study, Chowdhary and Srivastava (2013) reported a maximum total length of 32.5 cm and body weight of 251 g for magur collected from natural waters. The maximum total length and body weight reported for wild-caught magur by Kumar *et al*. (2017) was 28 cm and 119 g, respectively.

The variation in the growth of fish species is attributed to both genetic and non-genetic factors (Gjedrem, 2005). Similarly, the variation in growth of magur can be attributed to underlying genetic variation, then an optimally designed genetic selection program can ensure a sustainable supply of quality seeds. This will help in establishing commercial hatcheries of magur and the diversification of Indian aquaculture. Before starting any breeding program, it is essential to understand the amount of genetic variation present in the base population as it has a direct bearing on the response to selection. In animal breeding, genetic parameters are used to genetically evaluate a population for its genetic potential. The genetic parameters are estimated as the

functions of variance components obtained through implementing linear mixed models (LMM) on breeding data.

### 2.1 The general linear mixed model

Variance components and their functions are essential genetic parameters of interest in selection experiments. The linear mixed model (LMM) analysis, which is a popular tool for analysing breeding data, involves a two-step procedure where first, the variance components are estimated, followed by the prediction of breeding values assuming the estimated variance components are correct. The general form of a linear mixed model in a matrix format takes the following form (Mrode, 2014):

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{e} \quad \text{-----} \quad (2.1)$$

Where, $y = n \times 1$ vector of observations; $n =$ number of records; $b = p \times 1$ vector of fixed effects; $p =$ number of levels of fixed effects; $a = q \times 1$ vector of random family/animal effects; $a =$ number of levels of random effects; $e = n \times 1$ vector of random residual effects; $X =$ design matrix of order $n \times p$, which relates records to fixed effects; $Z =$ design matrix of order $n \times q$, which relates records to random genetic effects; $X$ and $Z$ are also known as incidence matrices.

According to Searle *et al.* (2009), it is important to recognize that a model is not just its equation given in 2.1, but a model is also everything that prescribes properties of the elements in that equation, which makes it essential to note the assumptions of the model. Model equation 2.1 has two parts which are termed fixed and random part. Since the model contains both the effects, it is known as mixed-effects model (Searle *et al*., 2009). The expectation of observation vector $\boldsymbol{E}(\boldsymbol{y}) = \boldsymbol{Xb}$ is the linear combination of fixed effetcs; random effects are assumed to have mean zero [$\boldsymbol{E}(\boldsymbol{a}) = \boldsymbol{E}(\boldsymbol{e}) = \boldsymbol{0}$], and they are independently and identically distributed (iid). Independent distribution refers to zero covariance between random effects levels, while identical distribution refers to homogeneity of variances across levels of random effects. The

random residual, $\boldsymbol{e} \sim \boldsymbol{iid}(\mathbf{0}, \boldsymbol{R})$, with $\boldsymbol{var}(\boldsymbol{e}) = \boldsymbol{R} = \boldsymbol{I}\sigma_e^2$, where $\sigma_e^2$ is the residual variance and $\boldsymbol{I}$ is the identity matrix; the random family/animal effects $\boldsymbol{a} \sim \boldsymbol{iid}(\mathbf{0}, \boldsymbol{G})$ where the structure of $G$ will depend on whether family or individual animal is assumed to be random. For random family effects: $\boldsymbol{var}(\boldsymbol{a}) = \boldsymbol{G} = \boldsymbol{I}\sigma_f^2$ where $\sigma_f^2$ is the family variance. For random animal effects: $\boldsymbol{var}(\boldsymbol{a}) = \boldsymbol{G} = \boldsymbol{A}\sigma_a^2$ where $\sigma_a^2$ is the additive genetic variance and $\boldsymbol{A}$ is the numerator relationship matrix which is given by twice the coefficient of coancestry (Falconer and Mackay, 1996).

### 2.2 Fixed and random effects

The effects included in LMM is classified as fixed and random effects. Different views and definitions are used under different contexts to classify an effect as fixed or random (Gelman, 2005). If the sample exhausts a population, the corresponding variable is fixed; whereas when the sample is a small part of the population the corresponding variable is random (Green and Tukey, 1960). . According to Robinson (1991) the fixed effects are estimated using least squares, and the random effects are estimated with shrinkage, which is also the standard definition in the multilevel modelling literature and in econometrics (Snijders and Bosker, 1999). The effects are fixed if they are constant across individuals, while random effects vary (Kreft and De Leeuw, 1998). The effects which are interesting in themselves are fixed, while if there is interest in an underlying population, they are random (Searle *et al.,* 1992). LaMotte (2014) explained that if an effect is assumed to be a realized value of a random variable, it is called a random effect.

The statistical criterion in the mixed model theory was used by Littell *et al.* (2006) to classify the fixed and random effects. Accordingly, a factor is random if its levels plausibly represent a larger population with a probability, whereas if an effect under study represents all possible levels of the factor or at least all levels about which inference is to be made, they are fixed effects. Schaeffer (2018) pointed, that if the estimated differences between the treatment effects would be expected to remain

consistent across multiple experiments, the treatment is assumed to be fixed. Whereas if it is possible to obtain a completely new set of factors when the experiments are repeated, such as sires, the factor can be assumed to be a random effect. Assuming an effect to be fixed, then the main objective is to estimate the treatment means or to test the treatment differences (Yang, 2010). Contrary, if the interest of inference would be in the variation of the effects of levels, then the factor is assumed as random (Schaeffer, 2018). Hence, in the same data, the effects can be switched between fixed and random depending on the goal of inference (Gelman, 2005). The suggestions are made as to how many levels a factor should have to estimate the variance-covariance parameters of random effects with precision. Stroup and Mulitze (1991) suggested that a factor should have more than ten levels before it can be considered random. In another report, James and Stein (1992) showed that BLUE could be beaten by BLUP when the effect in question has more than two levels.

The distinction between the fixed and random have many shades of grey, and hence the decision is made based on experience and tradition in a particular field of study (Schaeffer, 2019). Searle *et al*. (2009) noted different features to decide a set of factors as fixed or random which are the context of the data, the manner in which data were gathered and the environment from which they were obtained. In the animal breeding context, it is common to assume sires, dams and offspring as random effects. Other factors in the model can be interpreted as fixed or random based on one's interest in inferring the particular effect. Depending upon the treatment effects, the classification given by Eisenhart (1947) categorized a statistical model as fixed effects model if all the treatment effects are regarded as fixed effects, as random-effects model if all the effects are regarded as random and as mixed-effects model for some are regarded as fixed effects and some as random effects. In the context of animal breeding, mixed-effects model are of particular interest, for it is common to encounter both fixed and random effects in a breeding experiment.

## 2.3 Non-genetic factors

Any conceivable factor other than the genetic makeup of an individual, which can affect individuals' trait values, can be regarded as non-genetic factors. In aquaculture breeding programs, the commonly encountered non-genetic factors are age groups, sex, stock, year class or batch, pond, cage, farm, feed, etc. (Gjedrem, 2005). It becomes important to accurately quantify non-genetic factors to correct the data such that the animal's phenotype is a better predictor of its genotype. Also, the knowledge of the effect of non-genetic factors can influence breeding objective and selection methods (Gjedrem, 2012). Farias *et al.* (2017) explained that when the non-genetic factors like tank effect are large (common environmental effect), it reduces the efficacy of selection methods, especially the family selection, as it tends to reduce the prediction accuracy. Jousy *et al*. (2018) reported the effects of various non-genetic factors as age groups, culture type, pond and sex on the harvest body weight on magur in which they found all these factors to be significantly affecting the growth in magur.

One of the difficulties in the aquaculture breeding for genetic selection is to deal with the large amount of effects common to fullsibs resulting from separate rearing of families till the fish attain a size of tagging. If the common environmental effect persists until harvest, it will affect the estimates of variance components, inflating the heritability estimates. However, the following reports indicated that such effects are likely to diminish with age. In African catfish, *C. gariepinus*, reports suggest an insignificant effect of common environmental effect after six months of communal rearing (Srimai *et al.*, 2019). In channel catfish, the effects common to fullsibs were higher at stocking (19.2%), which decreased by harvest (11.3%) (Bosworth *et al.,* 2020). Further reports on rainbow trout (Elvingson and Johansson, 1993), tilapia (Rezk *et al.*, 2009), chinook salmon (Winkelman and Peterson, 1994), and gilthead seabream (Navarro *et al.*, 2009b) are available in which a decreasing pattern of common environmental effects was observed from stocking to harvest. Even though the effects common to fullsibs will eventually decrease towards harvest, not including them in the statistical models might result in

some of the variance associated with common environmental effects being attributed to additive genetic variance inflating heritability (Gallardo *et al.*, 2010; Gjerde *et al.*, 1994; Pante *et al.*, 2002). To the best of my knowledge, there are no reports on how long the common environmental effects (tank effects, maternal effects and any other effects common to fullsibs) persist in magur.

A common practice to start a genetic selection program is to establish a base population with a broad genetic base by collecting individuals from different stocks. It is important to quantify the effect of stock on growth traits to select the fish from the stock with better performance. Rameez *et al*. (2020) reported a significant difference between growth traits of magur collected from different stocks (Non-AP and AP stocks) where Non-AP stock was 13.5 g heavier than AP stock. In the study conducted by García-Celdrán *et al.* (2015) on gilthead seabream, the estimated effect of stock on growth and carcass traits were significant.

Occasionally, families are produced in batches across multiple year classes where batch becomes an important non-genetic factor. For instance, Saillant *et al.* (2006) conducted an experiment in sea bass where they produced families across two batches in which they found the significant effect of batch on the growth traits. If the differential growth across year classes is resulted from the difference in management practices, then there is a need to optimize the management practices across different year classes. Rameez *et al*. (2020) reported that the part of the significant difference between growth across the batch has resulted from the difference in body weight at stocking. The dynamicity of the aquaculture environment could significantly affect the production at harvest, possibly due to differential productivity of pond and management practices. Guan *et al.* (2016) estimated the effects of different non-genetic factors in which pond environment significantly affected the growth traits of turbot at harvest and also found a highly significant effect of body weight at stocking on the harvest traits.

In the context of mixed models, often, the non-genetic factors are quantified assuming that they are fixed effects and are estimated as best linear unbiased

estimation (BLUE) (Lynch and Walsh, 1998). Different methods have been derived for the estimation of fixed effect (Foulley, 2015). However, Isik *et al.* (2017) noted that, since, least squares means are easy to interpret and their direct relation to data, it is common to use them to quantify non-genetic factors.

### 2.4 Genetic parameter estimates

Heritability and genetic correlations are important genetic parameter estimates used to evaluate the genetic potential of breeding populations. Heritability estimates for body weight at harvest and other economically important traits in several aquaculture species have been well documented (Gjedrem, 1983; Gjedrem and Robinson, 2014). The following reports suggested a wide range (0 to 0.9) of heritability for different growth traits in aquaculture species. Jousy *et al.* (2018) reported a high heritability of 0.63 for harvest body weight in magur. In African catfish, high heritability (0.49 – 0.51) was reported for body weight, total length and standard length at stocking and moderate heritabilities ranging from 0.29 – 0.35 for the same traits at harvest and a low to moderate heritability for condition factor (0.06 – 0.31) (Srimai *et al*., 2019). A low heritability (0.05 to 0.15) for growth traits in African catfish was reported by Srimai *et al.* (2020). Bosworth *et al*. (2020) reported a low to moderate heritability for body weight at stocking (0.15) and harvest (0.21) in channel catfish, *Ictalurus punctatus*; further they reasoned that the low to moderate heritability is due to the presence of different stocks from various hatcheries in the breeding program. In striped catfish, *Pangasianodon hypophthalmus*, Vu *et al.* (2019) estimated medium to high heritability (0.31–0.54) for body traits, while they obtained a low heritability (0.17 to 0.20) for condition index. Li *et al.* (2019) estimated high heritabilities (0.61 – 0.81), for growth performance traits (body weight , total length, body height) and low heritability (0.11) for condition factor based on uni-trait and multi-trait linear mixed models in oliveflounder, *Paralichthys olivaceus*. In another study on turbot, *Scophthalmus maximus*, Guan *et al*. (2016) reported low to medium heritabilities of 0.16 to 34 for body weight, 0.17 to 0.34 for total length and very low heritability (0.01 – 0.04) for condition factor. Saillant *et al*. (2006) reported moderate

to high heritability of 0.21 to 0.56 for body weight and 0.18 to 0.40 for total length in sea bass. Moderate heritabilities of 0.28 to 0.34 and 0.27 to 0.35 for weight and length and a low heritability (0.13) for condition factor was reported in gilthead seabream (Navarro *et al.*, 2009a), and another study on the same species reported moderate heritability of 0.25 for both weight and length (García-Celdrán *et al*., 2015). In Japanese flounder, *P. olivaceus*, low to moderate heritability estimates ranging from 0.12 to 0.39 for body weight (BW), total length (TL), condition factor (K), and average daily gain (ADG) was observed (Li *et al.*, 2018).

The genetic and phenotypic correlations can vary between different traits. Various reports are available, showing the low to high genetic and phenotypic correlations spanning along positive and negative scales. Srimai *et al*. (2019) reported the genetic correlations between growth traits of African catfish, which were high (0.93 – 0.99); however, a low to a high genetic correlation between condition factor and other traits were observed at stocking (0.17 – 0.20) and harvest (0.55 – 0.72). They also reported a high phenotypic correlation between growth traits at both stocking (0.9 – 0.97) and harvest (0.88 – 0.98), whereas a low to moderate phenotypic correlations were estimated between growth traits and condition factor at stocking (0.09 – 0.16) and harvest (0.01 – 0.57). In striped catfish, the estimates of genetic and phenotypic correlations between body traits (weight and length) were high and close to one (Vu *et al*., 2019). Li *et al*. (2019) estimated positive and high (0.70 – 0.91) genetic and phenotypic correlations between body weight, total length and body height in oliverflounder. The genetic (0.99) and phenotypic (0.90) correlations between body weight and total length reported by Guan *et al*. (2016) in turbout were high, while a low positive (0.15) and negative (-0.23) phenotypic correlations were observed between body weight and condition factor as well as between total length and condition factor respectively. In Asian seabass, a high genetic (0.91 - 1) and phenotypic (0.91 – 0.95) correlations were observed between body weight and total length (Saillant *et al*., 2006). Navarro *et al.* (2009) and García-Celdrán *et al.* (2015) reported high and positive genetic (0.86 – 0.97) and phenotypic (0.82 – 0.93) correlations in gilthead seabream. In

Japanese flounder, (Li *et al*., 2018) reported high and positive (0.87–0.94) genetic and phenotypic correlations among BW, TL and ADG and estimated a low to moderate (−0.10–0.58) negative correlations between K and other three traits.

The above reports suggested different ranges for heritability and genetic correlations in different aquaculture species. The heritability indicates additive genetic variation, which can vary across the temporal and spatial scale, the model used for its estimation, the method of estimation, experimental design etc. (Walsh and Lynch, 2018). The reported genetic and phenotypic correlations were also spanned in different ranges. The different reported range of the magnitude of genetic parameters conveys that they are the properties of the population and environment in which they are cultured, and therefore, it is also important to know the magnitude of these parameters in the respective environments.

### 2.5 Different methods of estimation of variance components

#### 2.5.1 Analysis of Variance (ANOVA) method

The analysis of variance method is the oldest and widely used for estimating variance components. The estimators are obtained by equating the observed and expected mean squares and solving the resulting equations (Swallow and Searle, 1978). Under a balanced design, the ANOVA estimators are the best quadratic unbiased estimators (BQUE). Further, under the assumptions of normality, BQUE yield sampling variances (Searle, 1971; Searle and Gruber, 1971). The BQUE property holds good even without normality assumptions; however, under the added assumption of normality, the ANOVA estimators have the advantage of being uniformly the best-unbiased estimator (BUE) (Searle *et al*., 1992). However, the ANOVA method can sometimes yield negative variance estimates; also, even under normality, the ANOVA estimators have unknown distributions and sampling variances involving very complicated functions.

Further, under unbalanced data, the ANOVA methodology yields no optimum properties other than unbiasedness (Searle, 1995). For unbalanced data, Henderson (1953) suggested three ways, namely Henderson's method I, II, and III, for selecting a set of mean squares to equate to their expected values under unbalanced data. Searle (1995) discussed the limitations of the Henderson's method and observed that even when used outside these limitations, Henderson's methods yield estimators with no useful properties other than unbiasedness.

Hofer (1998) expounded the ANOVA method's inability to account for the covariances among effects of genetically related animals. The ANOVA method's lack of ability can yield biased estimates of variance components if the data are from the population under selection (Sorensen and Kennedy, 1984). Searle *et al*. (2009) criticized the merit of unbiasedness of ANOVA estimators of the variance components estimated from the animal breeding data. Their argument follows that the unbalanced and voluminous data generated from the breeding experiment does not go in hand with the underlying concept of unbiasedness, which says that under many repetitions of the same experiment, the mean of the estimator should yield the value of the parameter. However, in the breeding scenario, Searle *et al*. (2009) noted that it is impossible to get the idealized repetitions with the same degree of unbalancedness. In the light of the limitations posed by the ANOVA method, alternatives were sought, which naturally lead to the development of likelihood-based methods (Searle, 1995).

### 2.5.2 Likelihood-based methods

The method of maximum likelihood (ML) was introduced by Fisher (1922) and incorporated into variance component estimation by Hartley and Rao (1967). Unlike the ANOVA, the ML estimation assumes a multivariate normal distribution for the linear mixed models (Hofer, 1998). Conceptually, ML estimates the parameters of the distribution that maximize the likelihood of the observed data (Lynch and Walsh, 1998). The ML estimators have the properties of large sample, they are asymptotically unbiased and normally distributed with variance equal to the inverse of the expected (or Fisher)

information matrix, asymptotically consistent and efficient (Searle *et al*., 2009) . Hofer (1998) noted that the ML estimator is consistent in a sense, as the sample size approaches infinity, the ML estimator gives the parameter value i.e. error $(\hat{\theta} - \theta)$ converges to zero as $n \to \infty$. Asymptotically efficient means, as $n \to \infty$ the covariance matrix of the ML estimates achieves the Cramer-Rao lower bound of the covariance matrix of unbiased estimators (Casella and Berger, 1990). The ML estimator is deprived of the above features under limited sample size, where the variance estimates are biased under finite sample size (Lee, 2000).

The ML estimator takes no account of the loss in degrees of freedom that results from estimating the fixed effects (Foulley, 1993; Harville, 1977). The application of the ML method did not found attractive in animal breeding due to a large number of levels of fixed effects and a low number of observations per subclass. This circumstance lead Patterson and Thompson (1971) to modify the maximum likelihood to develop the Residual Maximum Likelihood (REML) method. The REML estimators maximize only the portion of a likelihood that does not depend on fixed effects; hence, REML is a restricted ML version (Lynch and Walsh, 1998). To account for the loss in degrees of freedom on estimating fixed effects, the REML method uses a linear transformation of vector of observations $y$, with the help of a transformation matrix $K$, so that REML maximizes $Ky$, such that $KX = 0$, where $X$ is the design matrix relating fixed effects with the observation vector (Lee, 2000). Both the REML and ML methods constrain the parameter space between zero and infinity. Hofer (1998) observed this constrain to be a source of bias, whereas others have touted this constrain as merit because the method will no longer yield negative variance estimates (Lynch and Walsh, 1998). Another advantage likelihood-based estimators have compared to ANOVA estimators is that ML/REML can account for selection bias resulting from its ability to integrate the pedigree information in the form of numerator relationship matrix (Robertson, 1977). In animal breeding, the data used to estimate variance components originates from selection experiments involving continuous culling of animals. The variance estimation using ANOVA is subjected to bias when data comes from selection experiments since the ANOVA

method assumes that data are randomly sampled. Under a balanced design, REML and ANOVA estimates are identical (Patterson and Thompson, 1975), which is not the case with ML estimates (Corbeil and Searle, 1976). REML has become the standard for variance component estimation in from animal breeding data due to the desirable properties.

#### 2.5.2.1 Computing algorithms for REML

The iterative algorithms to find REML estimates can be grouped according to the order of derivatives used (Harville and Callanan, 1990; Searle *et al*., 2009). They include direct search (derivative-free) methods using the first derivatives of likelihood, and methods using both first and second derivatives of the likelihood (Lee, 2000). Smith and Graser (1986) and Graser *et al.* (1987) described the derivative free method to solve REML, which Meyer (1989) implemented in DFREML suit and later in Wombat (Meyer, 2007). The most popular method involving the first derivative is Expectation-Maximization (EM) algorithm of Dempster *et al.* (1977), they are very slow in convergence but guaranteed to yield estimates within parameter space (Meyer, 1990). The commonly used second derivative algorithms are AIREML, Newton-Raphson (NR) algorithm, and Fisher scoring method (Gilmour *et al.*, 1995), and yield the asymptotic covariance matrix (inverse of the matrix of second partial derivatives also known as Hessain matrix) (Meyer and Houle, 2013). Among all the algorithms, the AIREML described by Gilmour *et al*. (1995) is most widely accepted as it is easier to compute than the alternative second derivative methods.

### 2.5.3 Sampling based methods

Parameter estimation based on the mixed models often rely on the Gaussian model's asymptotic results, for e.g. likelihood-based estimation, which may not be appropriate in some situations, especially under small sample size (Thai *et al.*, 2013a). When the appropriateness of using asymptotic results is questioned, sampling-based procedures can be used to estimate these quantities (Morris, 2002). The

sampling-based procedures give a full distribution of the estimates, which makes it convenient to study the sampling properties of the parameter estimates.

#### 2.5.3.1 Bootstrap methods

The bootstrap method represents an alternative approach for estimating a point estimate's precision and the confidence interval. The bootstrap method was first introduced for simple linear models for normally distributed observations with homoscedastic residual variance (Efron, 1979). The principal idea of bootstrap is to resample the observed data repeatedly to create datasets similar to the original dataset, then fit them to construct the distribution of an estimator or a statistic of interest (Efron, 1979; Efron and Tibshirani, 1994). The resampling should mimic the true data generating process to generate a bootstrap distribution close to the original sample's true distribution (El Halimi, 2009; Flachaire, 2005; Ocaña *et al.*, 2005). Hence, the classical bootstrap methods developed for simple linear models should be modified to consider the characteristics of mixed models (Das and Krishen, 1999). So, in the context of mixed-effects models using a numerator relationship matrix, the bootstrap should take into account the covariance of the individuals within a family and the individual residuals.

##### 2.5.3.1.1 Non-parametric and parametric bootstrap methods

The bootstrap method falls into two broad categories based on the assumptions as non-parametric bootstrap and parametric bootstrap. The non-parametric bootstrap method does not hold any distributional assumptions, in which the method generates new data by resampling from the empirical distribution of the sample (Thai *et al*., 2013). A convenient way to perform a non-parametric bootstrap based on the animal model is to resample from the predicted random animal effects and residuals to create new data sets, which Morris (2002) pronounced as BLUP based bootstrap, which while performing should account for the correlation structure of the data. Morris (2002) noted that even though bootstrapping the BLUPs seems like a natural extension of ' 'Efron's bootstrapping the residuals to mixed models, there is a possibility that the

procedure can underestimate the variation in the data. The BLUPs are shrinkage estimators and the amount of shrinkage to mean zero depends on the sample size and variance components (Searle *et al*., 2009). So there is a possibility that the resulting data sets generated from the BLUP-based bootstrapping could be less variable. The parametric bootstrap is based on strong distributional assumptions, such as a multivariate normal distribution that depends on the model and the distribution of parameters and error. In the context of an animal model, parametric bootstrap method can be used to sample values for all random effects fitted for the data and pedigree structure, from a multivariate normal distribution with a mean of zero and a given covariance matrix (Meyer, 2006).

#### 2.5.3.2 Jackknife method

The jackknife initially developed as a procedure for correcting bias of the parmeter estimated (Quenouille, 1949) was later used for constructing confidence limits for a large class of estimators (McIntosh, 2016). It is similar to the bootstrap in that it involves resampling, but instead of sampling with replacement, the method samples without replacement. Jackknife samples are selected by taking the original data vector and deleting one observation from the set (Miller, 1974). Roff (2008) used jackknife method to compare sire and dam heritability, in which one family was removed at a time to generate jackknife replicates. In a simulation study, Roff and Preziosi (1994) implemented the jackknife method to estimate the genetic correlation using the family mean method, where they could obtain valid estimates of generic correlations and standard errors. Extending the same principle to the animal model setting, jackknife estimate of variance components and heritability can be evaluated for each jackknife replicate. It is notable that for large samples the jackknife estimate of standard error is roughly equivalent to the delta method (McIntosh, 2016), which is used to approximate standard error of heritability obtained by the REML method.

#### 2.5.3.3 Asymptotic sampling

Meyer and Houle (2013) proposed an alternative method to evaluate asymptotic sampling characteristics by repeated sampling of parameter estimates of variance components from the asymptotic, multivariate normal distribution of likelihood estimates of variance components. The function(s) of interest like heritability can then be calculated for each sample by inspecting the replicates' distribution. The method offers a straightforward and computationally undemanding way to derive sampling distributions and confidence intervals for covariance components' estimates and functions (Meyer and Houle, 2013). The method of asymptotic sampling is implemented in the Wombat software (Meyer, 2007).

### 2.5.4 Bayesian method

Gianola and Fernando (1986), proposed the Bayesian approach as a strategy to solve problems like small sample size and negative estimates of variance components arising in the animal breeding theory. Blasco (2001) noted that the animal breeding adopted the Bayesian method as an alternative for a long-standing issue in the standard analysis using likelihood and BLUP, which according to is the uncertainty in the estimation of genetic variance and the prediction error variance of breeding values (Gianola *et al.,* 1994). The uncertainty surrounding the point estimates could have an exact account by arriving at the Bayesian marginal posterior distribution (Gianola *et al*., 1994). On the other end, when the data set is sufficiently large, the difference between the REML and Bayesian methods is merely philosophical, and the results are very similar (Robinson, 1991; Harville and Carriquiry, 1992). However, in majority studies, a large data set may not be available in many cases (Blasco, 2001).

The REML framework studies the distribution of estimator and expresses the uncertainty as the standard error of this distribution, whereas the Bayesian framework aims to obtain the probability density function of the parameter for a given set of data to obtain the most probable value of the parameter (Blasco, 2017). While it is

challenging to compute the marginal posterior distribution involving multiple integrals (Gianola and Rosa, 2015), improvement in computational efficiency has helped to overcome this difficulty, for instance, with Markov Chain Monte Carlo (MCMC) methods (Walsh, 2001). The MCMC methods simulate complex probability distributions (Smith and Roberts, 1993), the resulting marginal posterior distribution accounts for the uncertainty of the nuisance parameter by integrating over the distribution of nuisance parameter (Walsh and Lynch, 2018). Under Bayesian analysis, the posterior density of parameters is proportional to the information in the data in the form of likelihood and the prior density of the parameters (Walsh and Lynch, 2018). The same reasoning is used for prior sensitivity studies to infer the sufficiency of data (Blasco *et al.,* 1998). However, Blasco (2017) noted a caveat in employing prior sensitivity studies to infer the asymptotic sufficiency of data; he pointed out there is always a possibility that there exists a sharp prior which can overcome the effect of data; hence, one should take into account of this fact while inferring data based on multiple priors.

### 2.6 Prediction of breeding values

One of the main features distinguishing mixed model methodology from conventional linear model methods is the ability to estimate realized values of random variables, that is, specific random effects or linear functions of random effects (Stroup *et al.*, 2018). Generally, the function derived to obtain the realized value of a random effect is not called an estimator, as Henderson (1950) first did, as estimation applies to parameters, whereas, a random effect is a random variable it is called a predictor or prediction of random effect (Searle *et al*., 2009). The general problem here is to predict a value for an unobservable random variable (genetic merit or breeding value) from a vector of the observed random variable (body weight or any other measured traits), and it is assumed that both of these random variables are jointly distributed (Searle *et al*., 2009). Searle (1997) and Searle *et al*. (2009), explained why it is reasonable to consider the predictor of the realized value of a random variable as its conditional mean of the random variable. The property of conditional mean under the assumption of multivariate

normality has useful properties (Lynch and Walsh, 1998; Morrison *et al.*, 1976), which gives rise to the following simple expression of the predictor under one-way random model, as given by Searle *et al*. (2009).

$$\widehat{\boldsymbol{u}}_{\boldsymbol{i}} = \boldsymbol{E}(\widehat{\boldsymbol{u}}_{\boldsymbol{i}}|\overline{\boldsymbol{y}}_{\boldsymbol{i.}})$$

$$\widehat{\boldsymbol{u}}_{\boldsymbol{i}} = \frac{\boldsymbol{n_i\sigma_f^2}}{\boldsymbol{n_i\sigma_f^2 + \sigma_e^2}}(\overline{\boldsymbol{y}}_{\boldsymbol{i.}} - \boldsymbol{\mu})$$

Where ($n_i$) number of offspring per family and ($\sigma_f^2$- between family variance and $\sigma_e^2$- within family variance in the simple one-way random model. Small $n_i$ values shrink the prediction towards zero as compared to large $n_i$ values where the predictions are closer to $(\bar{y}_{i.} - \mu)$. The mixed model prediction of breeding value leads to what is known as the Best Linear Unbiased Prediction (BLUP) with the help of mixed model equations.

### 2.6.1 Best Linear Unbiased Prediction (BLUP)

Henderson (1949) developed mixed model equations (MME) through which fixed effects and breeding values can be simultaneously estimated. Henderson (1990) recalls the development of MME as resulted from his attempt to combine the power of least squares to estimate fixed effects, with the appealing features of selection index such as shrinkage, combining information from relatives, multiple trait evaluation, and optimizing the mean of selected individuals under certain restrictions. Henderson *et al.* (1959) proved that the solutions to fixed effects are Generalized Least Squares (GLS) estimates, and, in later work, Henderson (1963) gave the proof that the solution to the random effects obtained from the mixed model equations are the BLUP. The MME given by Henderson (1949) is:

$$\begin{bmatrix} \boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z} \\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}Z + G^{-1}} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'R^{-1}y} \\ \boldsymbol{Z'R^{-1}y} \end{bmatrix}$$

The BLUP is only a special case of predictors among the variety of different predictors of random effects. Peixoto and Harville (1986) had discussed the optimum properties of other predictors, including BLUP. Robinson (1991) gave an account of various derivations of BLUP and suggested that the one provided by Henderson (1950) is computationally the most efficient.

One of the attractive properties of the BLUP is its ability to reduce the prediction error variance by shrinking the predictions to their grand mean. Hill and Rosenberger (1985) showed that the shrinkage of breeding values toward the grand mean was greater when the difference between the least-squares estimates and the grand mean was large and the number of observations was small. The shrinkage property of BLUP makes adjustments for the outliers in tandem with the exceptionally high or low values in the data, especially when the number of observations is less. The property implied by the term "Best in BLUP is credited to its shrinkage property (Morris, 1983). Piepho *et al.* (2008) observed that, for the breeding value estimation, the bias incurred from shrinkage is more than offset by the reduction in variance, thus leading to a smaller mean squared error, which increases the prediction accuracy (Copas, 1983; Miller, 2002).

In practice, the BLUP obtained from the mixed model predictions is an empirical BLUP, meaning that the perdition equation's variance components are estimates and not the parameter values (Piepho *et al*., 2008). Often, the most critical practical issue is to obtain good estimates of variance components (Smith *et al.*, 2005), which depends foremost on sample size and the availability of complete data back to a base population, to which genetic variance components refer (Piepho and Möhring, 2006; Robin Thompson, 1973, 1979; Van der Werf and De Boer, 1990).

### 2.6.2 Breeding value prediction using pedigree information

A desirable feature of BLUP is its ability to borrow strength from relatives by exploiting genetic correlation arising from the pedigree. The closer the genetic

correlation with relatives, the more information can be extracted from their phenotypic information. The most common approach to exploiting pedigree information involves using the numerator relationship matrix (*A*) computed from the coefficient of coancestry (Mrode, 2014). The elements in the $A$ matrix is computed as twice the coefficient of coancestry, which is the probability that both alleles at a locus are identical by descent (Falconer and Mackay, 1996). The advantage of using the coefficient of coancestry in BLUP is that the breeding value can be estimated for each individual as opposed to a single value for the family. Another advantage of estimating breeding value using the coefficient of coancestry comes from the augmented information gained from exploiting the genetic correlation among relatives, which is achieved through embedding genetic correlation structure derived from quantitative genetic theory (Piepho *et al*., 2008). Henderson (1985) points out that, using the inverse of the numerator relationship matrix $A^{-1}$ in BLUP can effectively control bias in estimation breeding value that can be caused by selection. However, if complete records back to the base population are not included in the pedigree, it will result in biased estimates of variance components and there by BLUP (Sorensen and Kennedy, 1984; Schenkel *et al.*, 2002; Van der Werf, 2002). Goddard (1986) noted that the BLUP prediction of breeding value using the mixed model equation must include the whole selection history via $A$-matrix and the unbiased estimates of $G$ and $R$ from the base population. When the data used includes the base population and the $A$-matrix, the REML estimation of variance components will give unbiased estimates of $G$ and $R$ from the base population (Walsh and Lynch, 2018).

### 2.6.3 Breeding value predictions from multivariate model

The estimation of breeding value based on multi-trait evaluation was first introduced by Henderson and Quaas (1976), wherein they explained simultaneous evaluation of three traits in beef cattle. A multiple trait analysis involves the simultaneous evaluation of animals for two or more traits and uses the phenotypic and genetic correlations between the traits. Mrode (2014) noted multi-trait evaluation as the optimum methodology to evaluate animals as it increases the accuracy of evaluations. Further,

he mentions the gain in accuracy is proportional to the absolute difference between the genetic and residual correlations between the traits. One way to quantify the gain in predictions using multivariate analysis is the prediction error variances (PEV) which is the by-product of solving mixed model equations (Henderson, 1973). The accuracy of prediction increases as the PEV decrease. Schaeffer (1984) showed that the percentage reduction on the PEV depended on the difference between the error correlations and genetic correlations between the traits included in the multivariate model. He also noted that, when the absolute value of error correlation is less than the absolute value of genetic correlation, the trait with lower heritability achieves a greater percentage reduction of PEV and vice versa. In another study, Thompson and Meyer (1986) examined the case when two traits have equal heritability, they noticed an increase in accuracy of predictions as the genetic and environmental correlations are large and have opposite signs. Also, when the heritability of one trait is low, then an increase in accuracy is gained by including other trait/traits with high heritability in the model. Meyer (1983) examined the effect of data structure to improve the prediction accuracies, i.e., through more connections between the fixed effects and sires. In her work, it was shown that an improvement achieved in the data structure by including multi-traits is as important as extra genetic information gathered from multiple traits in increasing the accuracy of prediction. The more connections in data are a result of a non-zero residual covariance structure.

### 2.7 Analysis of longitudinal data

The measurements recorded at successive times or ages, such as body weight considered to be gradually changing, constitute the longitudinal data (Schaeffer, 2016). Such data are taken along some spatial scale, or any other continuous covariable, for e.g. age, referred to as control variable. In such a case, the criteria of growth is complete curves or trajectories rather than individual points. The interest is to quantify genetic values and their dispersion structure among records for a complete range of the control variable. Meyer and Kirkpatrick (2005) referred to this trait as function-valued

(FV) traits for that the corresponding curves are defined by mathematical functions, while Kirkpatrick *et al.* (1990) called such traits as infinite-dimensional traits, as they could, hypothetically, be measured infinitely many times along the continues scale of interest. In aquaculture species, the body weight or growth are measured on the same individual at different times or ages constituting a longitudinal data. In some instances, the multiple measurements of a trait from same individuals are treated as repeated records, but more generally measurements at different ages are considered to represent different traits (Meyer and Hill, 1997).

### 2.7.1 Multi-trait analysis of longitudinal data

Traditionally, repeated measurements are analyzed with repeatability models with the basic assumption of unit genetic correlation between repeated measurements as they are regarded as an expression of the same trait over time (Mrode, 2014). Meyer and Hill (1997) noted that the assumption of unit genetic correlation of repeatability models does not hold valid with repeated measurements of body weight. The next best alternative to repeatability models is unstructured multivariate models. Mrode (2014) has discussed in detail about implementing a multi-trait model with unstructured covariance matrix to account for both the genetic and permanent environmental correlations at different time points. Clearly, unstructured multivariate models are good to handle repeated measures, but only when few parameters are to be estimated. For an unstructured multivariate model, with $t$ distinct traits on $N$ individuals, there are $t(t+1)/2$ parameters to be estimated (Meyer and Hill, 1997). Often, estimating such humongous number of parameters in a full multivariate model with number of traits equal to number of ages would result in a highly overparameterized analysis. Meyer and Hill (1997) noted that to fit such an unstructured multivariate model in practical applications which require variance components to be estimated, 't' usually has to be small and N to be comparatively large for analyses to be feasible and estimates to be sufficiently accurate.

There were attempts to reduce the number of parameters fitted by incorporating specific covariance structure, viz., Wade *et al.* (1993) considered the application of an auto-regressive function to dairy cattle data. Meyer and Hill (1997) noted the requirement of an apriori knowledge about the number of independent traits represented by repeated measurements as a constrain in imposing a covariance structure. An eigenvalue decomposition or a canonical decomposition are the standard procedures to identify the number of independent combinations among the $t$ records (Mrode, 2014). Hayes and Hill (1980) applied such procedures to the genetic and phenotypic covariance matrices. The canonical decomposition has been used by Wiggans *et al.* (1996) to reduce the number of traits in a multivariate genetic evaluation model describing dairy production for milk, fat, and protein yields at ten individual days of lactation where they identified five canonical traits, thereby, drastically reducing the number of traits from 30 to 5.

### 2.7.2 Random regression models (RRM)

Random regression models (RRM) were i proposed as an alternative to overcome the problem of overparametrization in multiple-trait analyses (Henderson, 1982; Laird and Ware, 1982). Later, Schaeffer and Dekkers (1994) were the first to present the potential for the practical application of the RRM; the authors used the RRM to analyze the test day milk yield records in dairy cattle. According to Schaeffer (2016), RRM is used to model the trajectory of observations taken over time, referred to as a phenotypic fit, and to estimate dispersion parameters of regression coefficients that are the part of covariance functions. Numerous linear and nonlinear functions have been proposed for modelling the trajectory of effects included in the model (Oliveira *et al.*, 2019). The most commonly used functions are orthogonal polynomial functions, such as Legendre polynomials suggested by Kirkpatrick *et al*. (1990), and the smoothing polynomial functions such as spline (Wegman and Wright, 1983; White *et al.*, 1999) and B-spline (De Boor, 1980; Meyer, 2005). Oliveira *et al*. (2019) suggested considering each function's advantages and disadvantage before defining the function to be used to model

the trajectory. He noted that nonorthogonal polynomials could produce coefficients that are highly correlated, yielding computational problems during estimation or solving the MME. Thus, orthogonal polynomials have been preferred among the polynomial functions, because they give the lowest correlations among the estimated regression coefficients. Meyer (2005) noted that RRM based on Legendre polynomials tend to consider fewer coefficients than corresponding analysis using B-spline functions, which may allow using more parsimonious models. However, Misztal (2006) noted that analysis of RRM based on spline functions has the advantage of faster convergence over Legendre polynomials. Jamrozik *et al.* (2010) pointed out the important question while using spline functions are related to the number and location of knots (points at which independent linear segments are connected), which should be defined before the analysis. Schaeffer (2016) considered using spline functions to model the fixed regression part and Legendre polynomials to fit the random effects. There are instances where different orders of fit of Legendre polynomials are used for each random effect; however, Schaeffer (2016) suggested using the same order of Legendre polynomials for all random factors to reduce the computational complexity.

### 2.7.3 Covariance functions

The covariance functions (CF) were proposed through a series of work by Kirkpatrick and Heckman (1989) and Kirkpatrick *et al*. (1990). The fact that infinite-dimensional trait can take on a value at each of the infinite numbers of ages along the trajectory and could be represented by a continuous function. Kirkpatrick *et al*. (1990) proposed to use Legendre polynomials as the continuous function to model trajectories of a trait. . As the name suggests, the covariance function describes the covariance structure of an infinite-dimensional trait as a function of time. In that sense, the covariance function is the infinite-dimensional equivalent of a covariance matrix for a given number of records taken overtime at different ages (Mrode, 2014). When the number of traits, $t$, is less, then a full order fit is assumed, where the order of Legendre polynomials equal to $t$. But, then, as $t$ increases the number of parameters to be

estimated also increases. To avoid overparametrization, the order of fit is reduced to $k$ where $k < t$, such that the order of Legendre polynomials used to fit trajectory is $k$, which will estimate $k(k+1)/2$ parameters instead of $t(t+1)/2$. Thus, covariance functions can reduce the rank of the covariance matrix such that they are still equivalent to covariance matrices for traits with many longitudinal records (Oliveira *et al*., 2019). Meyer and Hill (1997) explained that a covariance function is merely the infinite-dimensional equivalent to a covariance matrix for a given number of records taken at different ages. Further, CF can give the covariances between any two records measured at given ages as a function of the ages and coefficients of Legendre polynomials. Meyer and Hill (1997) were the first to show the random regression model's equivalence to the covariance function. Van der Werf *et al.* (1998) and Mrode (2014) showed that if the same functions are used for e.g. Legendre polynomials to fit trajectories, then the models based on covariance functions are equivalent to RRM. The coefficients of CF can be either estimated by a weighted least square procedure as described by Kirkpatrick *et al*. (1990) or by a restricted maximum likelihood procedure as explained by Meyer and Hill (1997).

### 2.7.4 Eigenfunctions

Eigenfunctions are the infinite-dimensional equivalent of eigenvectors, for curves, and the eigenfunctions of genetic CF provide an insight into the expected transformation of trajectories when subject to selection (Meyer and Kirkpatrick, 2005). The procedure to yield eigenfunctions and eigenvalues of CF was described by Kirkpatrick and Heckman (1989). Meyer (2003) noted that eigenfunctions of genetic CF are of interest to animal breeders who want to predict how effective their selection of parents of the next generation will be and how selection at any point might affect the remainder of the curve. Van der Werf (2002) explained that the shape of the eigenfunctions and magnitude of the associated eigenvalues provide information about possible change of the growth curve and it can be utilized for optimizing selection of growth trajectories.

### 2.7.5 RRM in aquaculture

Implementation of RRM for genetic evaluation of aquaculture species is scarce, and only limited publications are available. One of the earlier studies in which RRM was used for genetic evaluation in rainbow trout was reported by McKay *et al.* (2002). They used Gibbs sampling to estimate the genetic and permanent environmental covariances between the coefficients of a cubic polynomial fit of growth trajectories of body weight records spanned over 54 different ages. Rutten *et al.* (2005) reported the use of RRM with covariance function in aquaculture species in Nile Tilapia. The authors used a quadratic Legendre polynomial to model both the fixed and random effects of the body weight trajectory recorded between 100 to 325 days. Turra *et al.* (2012) compared different RRM models based on the inclusion and exclusion of the family effects assuming both homogenous and heterogeneous residual variance structures for analyzing growth curves in Nile tilapia. In another study, Zhao *et al.* (2018) used quadratic Legendre polynomials to model growth in Japanese flounder. He *et al.* (2017) used a multivariate random regression model with Legendre polynomials of different orders to analyze genetic changes in growth traits recorded at 20 different time-point in GIFT tilapia. He *et* al. (2018) used RRM with Legendre polynomials in GIFT Tilapia to analyze body weight and five morphological traits.

# 3. MATERIALS AND METHODS

## 3.1 Brood selection and induced breeding

The animals used in the present study were the progeny from a base population formed by assembling the magur fish collected from the natural waters of Andhra Pradesh, Assam, and the West Bengal states of India. The brood fish were maintained at Fresh Water Fish Farm (FWFF), ICAR-Central Institute of Fisheries Education, Balabhadrapuram-Kakinada, Andhra Pradesh, India, as a part of an on-going genetic improvement programme of magur. The collected fish weighed 50 to 100 g and were transported by air and road to the FWFF. The fish were quarantined for two weeks and observed for diseases, deformities, and parasites. If any fish was found with disease and/or deformities they were removed from the stock. After 15 days of quarantine, the fish were tagged with Passive Integrated Transponder (PIT) tag. The male and female fish were identified and segregated and then reared separately in 200$m^2$ earthen ponds and fed twice a day with commercially available high protein content feed (minimum 30%) at the rate of 3% of the body weight twice per day. In the month of April and May, the ponds were completely drained, and magur fish were collected and examined for sexual maturity. The matured brooders were collected for induced breeding. The selected fish were maintained in separate ponds and fed with the same diet at the rate of 5% of body weight supplemented with Vitamin E and Zinc. The fish with about 150 g body weight were selected as brooders, and in addition, females were selected if they were oozing eggs upon applying gentle pressure on their abdomen. The breeding in magur coincides with the North-East monsoon period, and in the present study, the breeding period spanned from July-September.

A single dose of commercially available hormone Ovatide® (Hemmo Pharmaceuticals Pvt Ltd), a synthetic analogue of Gonadotropin-Releasing Hormone (GnRH) was used for induced breeding. An intramuscular injection of Ovatide® at the rate of 0.1 and 0.2 ml per 100 g body weight was administered to males and females to

induce the breeding. After the injection, male and female fish were kept in separate FRP tanks till the collection of gametes. The eggs were collected in plastic trays by stripping the females after 17 hours of Ovatide® injection. In males, the milt cannot be obtained through stripping. Therefore, the males were anaesthetized using clove oil at the rate of 2ppm and then sacrificed to collect testis. The sperm suspension was prepared by macerating the testis in normal saline solution (0.9% NaCl) in a mortar with a pestle. The sperms remain under the dormant condition in normal saline solution. Fertilization was induced by mixing the eggs collected in a plastic tray with sperm suspension and mixed thoroughly with water, which activates the sperm. The fertilized eggs were washed twice with water before being released into hatching units.

### 3.2 Production of families

#### 3.2.1 Mating design

A single pair mating design, where one male is mated with one female, was used to produce fullsib families. A total of 162 fullsib families were bred in two batches, and separate sets of brooders were used in both batches. In batch-1 (2014) 98 fullsib families were produced, and in batch-2 (2015) 64 fullsib families were produced. Out of 162 families, 100 full-sibs families could successfully hatch, and 78 fullsib families survived till tagging through harvest (39 each from batch-1 and batch-2).

#### 3.2.2 Larval rearing

The fertilized eggs from different fullsib families were incubated separately in an indoor hatching unit (1.25m X 0.45m X 0.2m) facilitated with flow-through systems for water exchange. The fertilized eggs were transferred into the hatching tubs by attaching them to the water hyacinth (Eichhornia) roots. Hatching of the fertilized eggs took place after 24-27 hours of fertilization, and the hatchlings got detached by themselves from the Eichhornia roots. The unfertilized eggs, which are opaque-white in color were still attached to the roots of Eichhornia, which made it easy to clean the

hatching tubs. After the complete yolk sac absorption (3 days post-hatch), larvae were fed *ad libitum* with live feed (Zooplankton dominated with *Moina)*. After ten days of hatching, early fry of 10-12 mm size was shifted to a rectangular indoor rearing unit of dimensions 3.0 X 0.6 X 0.45 m. During indoor rearing, the early fry was fed thrice a day in the morning, noon and evening with zooplankton collected from the specially prepared earthen pond. This was supplemented with crumbled starter feed (<250 microns) containing 35% protein at the rate of 3% body weight, followed by artemia flakes at the rate of 5% body weight, preferably in evening hours. The juveniles (fullsib families) were reared in the outdoor rearing units for 30 days and thereafter were transferred to separate cement tanks of dimensions (4.0 Lx 1.5 W x 1.0 m H) at a stocking density of 200 no per tank for family-wise rearing until they attained a tagging size of average 10-15 g.

### 3.2.3 Tagging and communal rearing

From each fullsib family, 30 fish were randomly selected for tagging with PIT tags. A total of 2328 fish were PIT-tagged from 78 fullsib families. Body weight and total length were recorded at the time of tagging. After tagging the fishes were kept under observation for 48 hours to check for mortality due to tagging stress. No significant mortality was observed within this period. If any mortality was observed; additional fishes were tagged to maintain the family size. After two days of tagging, the fishes were released into earthen ponds of 200 $m^2$ for communal rearing under monoculture and polyculture systems. The monoculture was practiced only in the first batch, where families were divided for monoculture and polyculture systems. In the monoculture system, 300 magur fish were stocked per pond. In the polyculture system, 200 magur fish were stocked per pond along with 100 rohu fingerlings.

### 3.3 Recording growth traits

Tagged magur fishes were cultured for 365 days under standard aquaculture conditions, and at the end of the 365 days, all the ponds were completely

drained, and the fish were handpicked for recording the traits. As magur has the habit of burrowing into the bottom mud, the complete harvest is only possible through handpicking. Body weight (BW), total length (TL), body depth (BD) and head width (HW) was recorded on each fish.

**Traits recorded at different time interval after tagging**

| Tagging | 3 months | 6 months | 9 months | 12 months |
|---|---|---|---|---|
| BW | BW | BW | BW | BW |
| TL | TL | TL | TL | TL |
| | BD | BD | BD | BD |
| | HW | HW | HW | HW |

Additionally, condition factor (k) and average daily gain were estimated as follows:

Condition factor is given by the following formula (Barnham, 2003).

$k = 10^2 (W/L^3)$ where,

k is the condition factor or coefficient of condition, W is the weight of fish in grams, L is the length of fish in centimetres.

The average daily gain is estimated as follows

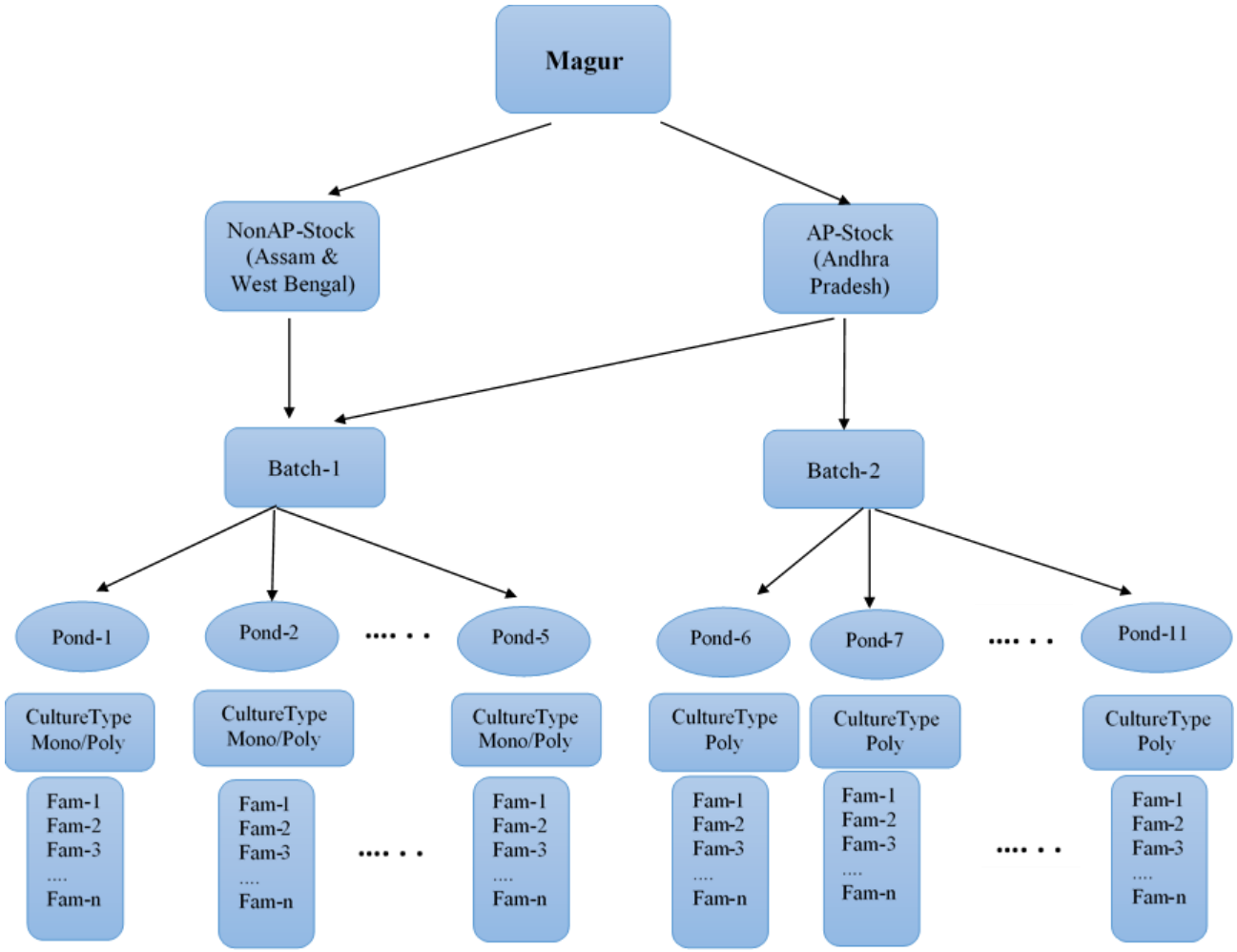
$ADG = (BW – BW_0)/D$

Where, BW is the body weight at harvest, $BW_0$ is the body weight at stocking, and D is the number of days, which is the culture period.

### 3.4 Statistical analysis

#### 3.4.1 Data structure

The following diagram illustrates the actual data generation process used in the present study.




#### 3.4.2 Normality testing

All traits were tested for normality (Kolmogorov-Smirnov test) using *Proc Univariate* in SAS® version 9.3. If data was found to be deviating from normality,

appropriate transformations were carried to bring the data to normal distribution. If any outliers were detected were removed from the data before further statistical analysis.

### 3.5 Quantification of the effect of non-genetic factors

#### 3.5.1 Non-genetic factors

The fish used in the present study were produced in two-year classes, 2014 and 2015. In 2014 Assam, Andhra Pradesh and West Bengal stocks were bred to produce the fullsib families. In 2015 there was no success in breeding from West Bengal and Assam stock. This resulted in the majority of the offspring coming from the Andhra Pradesh stock when data from both the year class were combined. To normalize the number of offspring per stock, Assam and West Bengal stocks were clubbed as Non-Andhra stock. In the present study, the batch (fish born in 2014 were considered as batch-1, and those born in 2015 were considered as batch-2), stock (Andhra and Non-Andhra stock), culture type (mono and polyculture), ponds (5 ponds from batch-1 and a separate set of 6 ponds from batch-2) and sex (F, M, and NA) were considered as non-genetic factors. The ponds were not the same for batch-1 and batch-2 but were on the same farm, and the same water source was used. The sex of those animals found dead before identifying them as female or male was marked NA, making three levels under the factor sex only for traits at stocking viz. M, F, and NA.

#### 3.5.2 Estimation of non-genetic factors

The non-genetic factors were quantified for traits at stocking and harvest. An ANOVA model with stock, batch, culture type, pond and sex as fixed effects and fullsib families as random effects was fitted with *proc mixed* in SAS 9.3 to test the significance of various non-genetic factors for harvest trait. Another model with stock, batch and sex as fixed effects and fullsib families as random effects was fitted with *proc MIXED* in SAS 9.3 to test the significance of various non-genetic factors for traits at stocking. For traits at harvest, body weight at stocking was used as a covariate in the

model. The least-squares means obtained after fitting the ANOVA model were used to quantify the effect of non-genetic factors. A multiple comparison test was made between different levels of each non-genetic factor using Tukey Kramer test. The following represents the ANOVA model used for the traits at stocking and harvest, respectively.

$$\boldsymbol{y_{ijknqi} = Stock_j + Batch_k + Sex_n + Family_{q(j,k)} + Error_i} \text{ --------}(1)$$

Where, $y_{ijknqi}$ is the trait recorded on $i^{th}$ individual at stocking, $Stock_j$ is the effect due to $j^{th}$ stock (j = 2), $Batch_k$ (K = 2)is the effect due to $k^{th}$ batch, $Sex_n$ (n = 3) is the effect of being male (M) or female (F) and NA, $Family_{q(jk)}$ (q = 78) is the effect of $q^{th}$ family produced from $j^{th}$ stock within $k^{th}$ $batch$.

$$\boldsymbol{y_{ijklmnqi} = BW0 + Stock_j + Batch_k + Culturetype_l + Pond_{m(k)} + Sex_n + Family_{q(jk)} + Error_i} \text{ ------- } (2)$$

Where, $y_{ijklmnqi}$ is the trait recorded on $i^{th}$ individual, $BW0$ is the initial body weight fitted as a covariate, $Stock_j$ is the effect due to $j^{th}$ stock, $Batch_k$ is the effect due to $k^{th}$ batch, $Culturetype_l$ is the effect due to $l^{th}$ culture type, $Pond_{m(k)}$ is the effect of $i^{th}$ pond belonging to $k^{th}$ batch, $Sex_n$ is the effect of being male (M) or female (F) in case of harvest traits and M, F or NA for traits at stocking , $Family_{q(jk)}$ is the effect of $q^{th}$ family produced from $j^{th}$ stock with in $k^{th}$ $batch$.

### 3.6 Estimation of variance components and heritability ($h^2$)

Variance components and heritability were estimated for the traits at stocking and harvest. Variance components and heritability estimates of all the traits were estimated by the Residual Maximum Likelihood (REML) method.

The variance components and heritability of body weight at harvest were estimated employing different methods/models. The following methods were used:

1. Analysis of Variance (ANOVA) method
2. Likelihood-based methods
    a. Maximum Likelihood method (ML)
    b. Residual Maximum Likelihood method (REML)
3. Resampling based methods
    a. Non-parametric bootstrap
    b. Parametric bootstrap
    c. Jackknife method
    d. Asymptotic sampling
4. Multivariate approach using REML method
5. Bayesian Estimation

### 3.6.1 Statistical models for variance estimation

The statistical analysis was performed under the framework of a linear mixed model (LMM). Two different variations of the LMM were used, viz a family model and/or an animal model depending on the method used. For instance, the traditional ANOVA model based on the method of fitting constants cannot make use of the pedigree information, so a family model was used. Whereas an animal model was employed when the likelihood based methods were used for estimation. In matrix notation, an LMM takes the following form

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e} \text{ ---------} (3)$$

Where $\boldsymbol{y}$ is $\boldsymbol{n} \times \boldsymbol{1}$ random vector of observations, $\boldsymbol{b}$ is $\boldsymbol{p} \times \boldsymbol{1}$ vector of fixed effects, $\boldsymbol{u}$ is $\boldsymbol{q} \times \boldsymbol{1}$ the vector of random effects, and $\boldsymbol{e}$ is $\boldsymbol{n} \times \boldsymbol{1}$ column vector of random residual effects. The $\boldsymbol{X}$ and $\boldsymbol{Z}$ are respectively $\boldsymbol{n} \times \boldsymbol{p}$ and $\boldsymbol{n} \times \boldsymbol{q}$ design matrices relating observations to the levels of fixed effects and random effects, respectively. The assumptions regarding random vector $\boldsymbol{y}$ were as follows:

$$\boldsymbol{y} \sim \boldsymbol{MVN}(\boldsymbol{Xb}, \boldsymbol{ZGZ'} + \boldsymbol{R}) \text{-------} (4)$$

$$\boldsymbol{E(y) = Xb}\ and\ \boldsymbol{Var(y) = ZGZ' + R}\ \text{--------}(5)$$

Where $\boldsymbol{E(y)}$ is the expectation or mean vector of $\boldsymbol{y}$ ; $\boldsymbol{Var(y)}$ is the variance of $\boldsymbol{y}$ .

The fixed effect vector remained the same for both the family and animal models: BW0 (covariate), stock, batch, culture type, pond, and sex. Whereas the random effects of interest in a family model were fullsib families and animal model individual animals were the random effects. The following assumptions were made for each model regarding random effects. It was assumed that random effects ($\boldsymbol{u}\ and\ \boldsymbol{e}$) followed a multivariate normal distribution.

$$\mathbf{u} \sim \mathbf{MVN}(\mathbf{0}, \boldsymbol{G})\ and\ \mathbf{e} \sim \mathbf{MVN}(\mathbf{0}, \boldsymbol{R})\ \text{---------}\ (6)$$

By definition:

$$\boldsymbol{E(u) = E(e) = 0}\ \text{---------}\ (7)$$

$$\boldsymbol{Var(u) = G}\ and\ \boldsymbol{Var(e) = R}\ \text{---------}\ (8)$$

Where, $\boldsymbol{E(u)}\ and\ \boldsymbol{E(e)}$ were the expectation of random genetic effects and residual effects, $\boldsymbol{G}$ and $\boldsymbol{R}$ were the variance-covariance matrices of random genetic effects and residual effects, respectively.

### 3.6.1.1 Fullsib model

The fullsib model has the following assumptions regarding $\boldsymbol{G}$ and $\boldsymbol{R}$:

$$\boldsymbol{G} = \mathbf{I}\boldsymbol{\sigma_f^2}\ \text{---------}\ (9)$$

Where, $\boldsymbol{\sigma_f^2}$ is the between family variance and $\boldsymbol{I}$ the identity matrix;

$$\boldsymbol{R} = \boldsymbol{I}\boldsymbol{\sigma_e^2}\ \text{---------}\ (10)$$

Where, $\boldsymbol{\sigma_e^2}$ is the residual variance and $\mathbf{I}$ is the identity matrix.

The identity matrix in the function of $\boldsymbol{G}$ and $\boldsymbol{R}$ followed from the assumption of statistical independence. The additive genetic variance $\boldsymbol{\sigma_a^2}$ from the fullsib model was obtained as twice the family variance $\boldsymbol{\sigma_f^2}$

$$\boldsymbol{\sigma_a^2 = 2 \times \sigma_f^2} \text{ ---------} (11)$$

### 3.6.1.2 Animal model

The assumptions regarding $\boldsymbol{G}$ and $\boldsymbol{R}$ matrices for the animal model was as follows:

$$\boldsymbol{G = A\sigma_a^2} \text{ ---------} (12)$$

Where, $\boldsymbol{\sigma_a^2}$ is the additive genetic variance and $\boldsymbol{A}$ is the additive genetic (or numerator) relationship matrix and has elements

$$\boldsymbol{A_{ij} = 2\Theta_{ij}} \text{ ---------} (13)$$

Where, $\boldsymbol{\Theta_{ij}}$ is the coefficient of coancestry, which is the probability that an allele drawn at random from individual $\boldsymbol{i}$ will be identical by descent to an allele drawn at random from individual $\boldsymbol{j}$.

$$\boldsymbol{R = I\sigma_e^2} \text{ ---------} (14)$$

Where, $\boldsymbol{\sigma_e^2}$ is the residual variance and $\mathbf{I}$ is the identity matrix.

In addition to model-1 under the animal model, another model was used where the pond effect was also assumed to be a random effect in addition to the animal effect. It is presented in the matrix notation as follows:

$$\boldsymbol{y = Xb + Zu + Wp_o + e} \text{ ---------} (15)$$

Where, $\boldsymbol{y}$, $\boldsymbol{b}$, $\boldsymbol{u}$, $\boldsymbol{e}$, $\boldsymbol{X}$ **and** $\boldsymbol{Z}$ and their assumptions are the same as explained previously. The newly introduced terms $\boldsymbol{W}$ is the design matrix relating observations to the levels of random pond effects and $\boldsymbol{p_o}$ is the vector of random pond effects having the following assumptions:

$$\boldsymbol{p_o} \sim \mathbf{MVN}\left(\mathbf{0}, \mathbf{I}\boldsymbol{\sigma^2_{p_o}}\right) \text{ ---------} \ (16)$$

Where, $\boldsymbol{\sigma^2_{p_o}}$ is the pond variance and $\mathbf{I}$ is the identity matrix. Model-2 was used only for the analysis of harvest traits under the REML method.

The heritability $\boldsymbol{h^2}$ of each trait was calculated as:

From model represented in equation $(\mathbf{3})$

$$\boldsymbol{h^2} = \frac{\boldsymbol{\sigma^2_a}}{\boldsymbol{\sigma^2_a + \sigma^2_e}} \text{ ---------} \ (\mathbf{17})$$

From model represented in equation $(\mathbf{15})$

$$\boldsymbol{h^2} = \frac{\boldsymbol{\sigma^2_a}}{\boldsymbol{\sigma^2_a + \sigma^2_{p_o} + \sigma^2_e}} \text{ ---------} \ (18)$$

The models represented by equations $(3)$ and $(15)$ were compared by means of the likelihood ratio test as follows (Littell *et al*., 2006).

$$\boldsymbol{H_0}: \boldsymbol{\sigma^2_g} = \mathbf{0} \text{ against } \boldsymbol{H_1}: \boldsymbol{\sigma^2_g} > \mathbf{0}$$

Test Statistic

$$\boldsymbol{D} = \mathbf{2}[\boldsymbol{logL_F} - \boldsymbol{logL_R}]$$

Where $\boldsymbol{D}$ is twice the difference between full model REML log-likelihood $(\boldsymbol{logL_F})$ and reduced model REML log-likelihood $(\boldsymbol{logL_R})$. The values of $\boldsymbol{D}$ are

distributed approximately as Chi-square with degrees of freedom equal to the difference between the number of parameters of the models.

### 3.6.2 Analysis of Variance (ANOVA) method

The ANOVA method was implemented by employing the fullsib model in *proc glm* in SAS 9.3. The family variance was estimated by equating sums of squares involving random effects to their expectations (Searle, 1995).

$$\boldsymbol{MSS_f = K.\sigma_f^2 + \sigma_e^2} \text{ ---------} (19)$$

Family variance was be obtained as:

$$\boldsymbol{\sigma_f^2 = \frac{MSS_f - \sigma_e^2}{K}} \text{ ---------} (20)$$

Additive genetic variance:

$$\boldsymbol{\sigma_a^2 = 2 \times \sigma_f^2} \text{ ---------} (21)$$

Phenotypic variance:

$$\boldsymbol{\sigma_p^2 = \sigma_f^2 + \sigma_e^2} \text{ ---------} (22)$$

Where, $\boldsymbol{MSS_f}$ is the mean square of family effect obtained from *proc glm,* $\boldsymbol{K}$ is the weighted average number of offspring per family, $\boldsymbol{\sigma_f^2}$ is the family variance, $\boldsymbol{and\ \sigma_e^2}$ is the residual variance (obtained as direct output from *proc glm* as error mean square). The heritability was estimated as

$$\boldsymbol{h^2 = \frac{\sigma_a^2}{\sigma_p^2}} \text{ ---------} (23)$$

The standard error of heritability was obtained as follows (Becker, 1992):

$$SE(h^2) = 2\sqrt{\frac{2\,(n-1)\,(1-t)^2\,[1+(k-1)\,t]^2}{k^2\,(n-f)\,(f-1)}} \text{ --------- } (24)$$

Where, $f$ is the number of fullsib families, $n$ is the total number of individuals, $k$ is the weighted average number of offsprings per family and $t$ is the intraclass correlation given by:

$$t = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_e^2} \text{ --------- } (25)$$

### 3.6.3 Likelihood based methods

#### 3.6.3.1 Maximum likelihood (ML) method

ML method was implemented under *proc mixed* in SAS 9.3 employing, both the fullsib and animal model (Littell *et al*., 2006). Under ML fullsib model $\sigma_f^2$ and $\sigma_e^2$ were directly obtained.

Additive genetic variance:

$$\sigma_a^2 = 2 \times \sigma_f^2 \text{ --------- } (26)$$

Phenotypic variance:

$$\sigma_p^2 = \sigma_f^2 + \sigma_e^2 \text{ --------- } (27)$$

Under ML animal model $\sigma_a^2$ and $\sigma_e^2$ are directly obtained. In both cases, heritability was estimated as:

$$h^2 = \frac{\sigma_a^2}{\sigma_p^2} \text{ --------- } (28)$$

The sampling variance of the variance parameters was obtained from *proc mixed*, whereas the standard error of heritability is estimated by implementing delta

method in *proc IML* as detailed by Isik, (2009). The SE of heritability was obtained by the following approximation:

$$\boldsymbol{SE(h^2)} = \sqrt{\left(\frac{2\sigma_f^2}{\sigma_p^2}\right)^2 \left[\frac{Var\left(2\sigma_f^2\right)}{\left(2\sigma_f^2\right)^2} + \frac{Var(\sigma_p^2)}{(\sigma_p^2)^2} - \frac{2Cov\left(2\sigma_f^2,\sigma_p^2\right)}{2\sigma_f^2\sigma_p^2}\right]} \text{ --------- } (29)$$

Where,

$\boldsymbol{Var\left(2\sigma_f^2\right)}$ is the variance of the family variance, which was obtained from the output table called *Asymptotic Covariance Matrix of Estimates* of SAS proc mixed procedure, $\boldsymbol{Var(\sigma_p^2)}$ is the variance of phenotypic variance:

$$\boldsymbol{Var(\sigma_p^2) = Var(\sigma_p^2 + \sigma_e^2) = Var(\sigma_f^2) + Var(\sigma_e^2) + 2Cov(\sigma_f^2, \sigma_e^2)} \text{ --------- } (30)$$

The above expression was obtained by summing all the variances of variance components plus two times the covariance between family variance and error variance components, all of which were directly obtained from the output table called Asymptotic Covariance Matrix of Estimates of SAS proc mixed procedure.

$\boldsymbol{Cov\left(2\sigma_f^2, \sigma_p^2\right)}$ is the covariance between family variance and phenotypic variance, which expands as follows:

$$\boldsymbol{Cov\left(2\sigma_f^2, \sigma_p^2\right) = Cov\left(2\sigma_f^2, \left(\sigma_f^2 + \sigma_e^2\right)\right)}$$

$$\boldsymbol{= Cov\left(2\sigma_f^2, \sigma_f^2\right) + Cov\left(2\sigma_f^2, \sigma_e^2\right)}$$

$$\boldsymbol{= 2Var\sigma_f^2 + 2Cov\left(\sigma_f^2, \sigma_e^2\right)} \text{ --------- } (31)$$

Which was obtained from the output table called *Asymptotic Covariance Matrix of Estimates* of SAS proc mixed procedure.

### 3.6.3.2 Residual Maximum likelihood (REML) method

REML fullsib model was implemented (Littell *et al*., 2006) in *proc mixed* in SAS 9.3 to obtain $\boldsymbol{\sigma_f^2}$ and $\boldsymbol{\sigma_e^2}$ to estimate following variance parameters.

Additive genetic variance:

$$\boldsymbol{\sigma_a^2 = 2 \times \sigma_f^2} \text{ ---------} \quad (32)$$

Phenotypic variance:

$$\boldsymbol{\sigma_p^2 = \sigma_f^2 + \sigma_e^2} \text{ ---------} \quad (33)$$

Heritability:

$$\boldsymbol{h^2 = \frac{\sigma_a^2}{\sigma_p^2}} \text{ ---------} \quad (34)$$

REML animal model was implemented in Wombat (Meyer, 2007). Wombat directly output the values of $\boldsymbol{\sigma_a^2}$ , $\boldsymbol{\sigma_e^2}$ and $\boldsymbol{\sigma_p^2}$ along with their sampling variance, further it gave the estimates of $\boldsymbol{h^2}$ and its standard error.

### 3.6.4 Resampling based methods

### 3.6.4.1 Nonparametric bootstrap

Nonparametric bootstrapping (Efron, 1992) was performed under two different schemes, namely global bootstrap and subject wise bootstrap (Thai *et al.*, 2013), the major difference between them being the basic resampling unit.

#### 3.6.4.1.1 Nonparametric global (random and residual effects) bootstrap

In global bootstrap, samples were drawn after modifying Thai *et al*. (2013) from the predicted random animal effects and residuals where the resampling unit was

the whole vector of predicted random animal and residual effects. The family structure of the data was not considered for resampling under this method.

#### 3.6.4.1.2 Nonparametric subject wise (random and residual effects) bootstrap

In subject wise bootstrap, samples were drawn from the predicted random animal effects and residuals where the resampling unit was a parent and its offspring (Thai *et al*., 2013), taking into account the family structure of the data.

A total of 10,000 new data sets were generated, and for each data set, heritability was estimated using the REML procedure to obtain the sampling distribution of heritability estimates. The bootstrap samples were obtained by implementing the following algorithm:

1. Fit the model, $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$, to the data
2. Predict the random animal effects $\hat{\boldsymbol{u}}$ and residuals $\hat{\boldsymbol{e}}$
3. Obtain vector $\hat{\boldsymbol{y}} = \boldsymbol{y} - (\hat{\boldsymbol{u}} + \hat{\boldsymbol{e}})$, where $\hat{y}$ is $y$ vector corrected for effects of $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{e}}$
4. Draw random animal effects with replacement from predicted $\hat{\boldsymbol{u}}$ (step 2) to obtain new $\hat{\boldsymbol{u}}^*$
5. Draw residuals with replacement from predicted $\hat{\boldsymbol{e}}$ (step 2) to obtain new $\hat{\boldsymbol{e}}^*$
6. Create new $\boldsymbol{y}$ values $\hat{\boldsymbol{y}}^*$ as $\hat{\boldsymbol{y}}^* = \hat{\boldsymbol{y}} + \hat{\boldsymbol{u}}^* + \hat{\boldsymbol{e}}^*$
7. Fit the model $\hat{\boldsymbol{y}}^* = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$ to the new data set to estimate variance components and obtain heritability
8. Repeat the steps from 4 to 7 for 10,000 times

The above algorithm was implemented in R software (version 4.0.0), where part of the task (resampling) was performed within the R environment, and the variance estimation was performed in Wombat (Meyer, 2007). After every resampling, the new bootstrap data set was pipelined into the Wombat (by calling Wombat from R

environment), to estimate variance parameters and heritability. The variance estimation was performed by REML method in Wombat (Meyer, 2007).

### 3.6.4.2 Parametric bootstrap

In parametric bootstrapping, the *new* data sets were generated by sampling the values for random animal and residual effects from a multi-variate normal distribution (Thai *et al*., 2013), with a mean of zero and covariance matrix as estimated from the REML analysis. The Parametric bootstrap sample was obtained as follows:

1. Fit the model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$ to the data
2. Estimate the overall mean $\boldsymbol{E(y)} = \boldsymbol{X\beta}$
3. Simulate the random animal effect $\hat{\boldsymbol{u}}$ and residuals $\hat{\boldsymbol{e}}$ from the multivariate normal distribution
4. Create new $\hat{\boldsymbol{y}}$ values as $\hat{\boldsymbol{y}} = \boldsymbol{E(y)} + \hat{\boldsymbol{u}} + \hat{\boldsymbol{e}}$
5. Fit the model $\hat{\boldsymbol{y}} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$ to the new data set to estimate variance components and obtain heritability
6. Repeat steps 3 to 5 for 10,000 times

Sampling from multivariate normal distribution and estimation of variance components was implemented within Wombat (Meyer, 2007). For simulating random effects in Wombat (both animal and residual effects), a built-in function (runop --*simul*) was used. Since the number of simulated data set that could be generated using the built-in function was limited by 999, an algorithm was implemented in R (Version 4.0.0) software to make Wombat more flexible to generate 10,000 simulated data sets.

### 3.6.4.3 Jackknife (Leave-One-Out) method

Jackknife samples were generated according to McIntos (2016) from the data by leave-one-out method. The original data had 1413 observations from which 1413 unique jackknife samples were generated by leaving out one observation per every

jackknife sample. The following algorithm was implemented to generate jackknife estimates:

For $\boldsymbol{i = 1\ to\ 1413}$ repeat the following:

1. Leave out the $\boldsymbol{i^{th}}$ observation to form the $\boldsymbol{i^{th}}$ jackknife sample
2. Fit the model $\widehat{\boldsymbol{y}} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$ to the $\boldsymbol{i^{th}}$ jackknife sample to estimate variance parameter and heritability.

The algorithm was implemented in R (version 4.0.0) software to generate jackknife samples and then pipelined those into Wombat (Meyer, 2007) to estimate variance parameters and heritability.

### 3.6.5 Asymptotic sampling

In asymptotic sampling (Meyer and Houle, 2013), 10,000 samples of variance parameters were generated by sampling from an asymptotic distribution of likelihood estimates. Sampling was done in the following fashion:

1. Fit the model $\widehat{\boldsymbol{y}} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$ to estimate the variance components and corresponding asymptotic multivariate normal distribution.
2. Sample variance components (additive genetic and residual variances) from the distribution.

Asymptotic sampling was performed in Wombat (Meyer, 2007) using built-in function *--sample*.

### 3.6.6 Bayesian estimation

#### 3.6.6.1 Prior distribution

An inverse-gamma distribution was used as the prior for variance components, and a diffuse normal prior centered around zero with a very large variance

($10^8$) was used for the mean (Hadfield *et al*., 2019). The inverse-gamma distribution was parameterized by two parameters nu (shape) and V (scale), where nu = 0.002 and V = 1.

### 3.6.6.2 Markov Chain Monte Carlo (MCMC) sampling

MCMC sampling was implemented in the R software with the help of *MCMCglmm* package (de Villemereuil, 2012). The *MCMCglmm* package used a combination of Gibbs sampling, slice sampling, and Metropolis-Hastings updates to generate the chain (Hadfield *et al*., 2019). The starting value was generated by default heuristic techniques within *MCMCglmm,* which was then used to initialize the chain. A single chain was generated with a chain length of $\mathbf{10,00,000}$ iterations with a thinning interval of $\mathbf{100}$ iterations (only one iteration value is saved at an interval of $\mathbf{100}$ iterations) to reduce the autocorrelation between successive values (de Villemereuil, 2012). Before saving iteration values, a burn-in period of $\mathbf{50,000}$ iterations was ensured for better convergence of the chain.

### 3.6.6.3 Test of convergence

Convergence was first assessed visually by looking at time series trace plots of parameters. Geweke's statistic (Geweke, 1992) was estimated as a statistical test for convergence using *gewke.diag()* function from *coda* package (Plummer *et al.*, 2006) in the R (version 4.0.0). The test statistic was a standard Z-score based on a test for equality of the means of the first and last part of a Markov chain (by default, the first 10%, and the last 50%).

### 3.6.6.4 Highest density regions (HDRs)

HDR, also known as Bayesian confidence intervals or credible intervals, was used to summarize the posterior distribution. A 95% HDR was interpreted as a $(1 - \alpha)$ (0.95) probability that the interval contains the true value of the unknown parameter. Whereas $1 - \alpha$ (0.95) probability from frequentist methods implies that 95% of the time

the confidence intervals will enclose the parameter. HDRs were estimated by the function *HPDinterval()* from *coda* package (Plummer *et al*., 2006) in R (version 4.0.0).

Three measures of the location were used to summarise the posterior densities viz the mean, median, and mode. The mean and median were estimated using standard methods. The mode of posterior distribution was obtained by using the function *posterior.mode()* from *coda* package in R (Version 4.0.0) software. The kernel density plots were obtained by using *proc kde* in SAS 9.3.

### 3.7 Prediction of breeding values (BVs)

The BVs were predicted as the Best Linear Unbiased Predictors (BLUPs) under the framework of LMMs using an animal model.

#### 3.7.1 Univariate Best Linear Unbiased Predictor (uBLUP)

The BLUP BVs were predicted using the following LMM model:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e} \text{ ---------} \quad (37)$$

The component vectors and matrices of the above equation was the same as explained under the context of an animal model (section 3.6.1, equation$(3)$). The BVs were obtained by solving Mixed Model Equations (MME), which involves the model component vectors and matrices $\boldsymbol{y}$ , $\boldsymbol{X}$, $\boldsymbol{Z}$, $\boldsymbol{G}$, and $\boldsymbol{R}$. For the above mixed model, $\boldsymbol{y}$, $\boldsymbol{X}$, and $\boldsymbol{Z}$ were already observed from the data, while $\boldsymbol{G}$ and $\boldsymbol{R}$ were estimated using the animal model. The general form of MME is as follows:

$$\begin{bmatrix} \boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z} \\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}Z + G^{-1}} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'R^{-1}y} \\ \boldsymbol{Z'R^{-1}y} \end{bmatrix} \text{ ---------} \quad (38)$$

Assuming $\boldsymbol{R}$ to be identical and independent for all observations, it can be factored out from both sides of the equation. The general form of MME, when used under

the context of the animal model, was reduced to the following expression (Henderson, 1973) :

$$\begin{bmatrix} \boldsymbol{X'X} & \boldsymbol{X'Z} \\ \boldsymbol{Z'X} & \boldsymbol{Z'Z + \lambda A^{-1}} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'y} \\ \boldsymbol{Z'y} \end{bmatrix} \text{---------} (39)$$

Where,

$$\boldsymbol{\lambda} = \frac{\boldsymbol{\sigma_e^2}}{\boldsymbol{\sigma_a^2}} = \frac{\boldsymbol{(1-h^2)}}{\boldsymbol{h^2}} \text{---------} (40)$$

Where, $\boldsymbol{\sigma_e^2}$ is the residual variance, $\boldsymbol{\sigma_a^2}$ the additive genetic variance and $\boldsymbol{h^2}$ the heritability.

The univariate BLUP BVs ($\widehat{\boldsymbol{u}}$**)** were obtained by solving the above equation to obtain:

$$\widehat{\boldsymbol{u}} = [\boldsymbol{Z'X} \quad \boldsymbol{Z'Z + \lambda A^{-1}}]^{-1} \begin{bmatrix} \boldsymbol{X'y} \\ \boldsymbol{Z'y} \end{bmatrix} \text{---------} (41)$$

### 3.7.2 Accuracy of univariate breeding values

The standard error of BLUP breeding values was estimated as the square root of Prediction Error Variance (PEV), and is obtained from the diagonal elements of the inverse of MME (Henderson, 1975). The generalized inverse of the Left Hand Side (LHS) matrix of MME can be represented as:

$$\begin{pmatrix} \boldsymbol{X'X} & \boldsymbol{X'Z} \\ \boldsymbol{Z'X} & \boldsymbol{Z'Z + \lambda A^{-1}} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{C^{11}} & \boldsymbol{C^{12}} \\ \boldsymbol{C^{21}} & \boldsymbol{C^{22}} \end{pmatrix} \text{---------} (42)$$

Then, the prediction error variance is given by the following:

$$\boldsymbol{PEV} = \boldsymbol{C^{22}\sigma_e^2} \text{---------} (43)$$

The PEV for the BV of $i^{th}$ the animal was obtained as:

$$\boldsymbol{PEV_i} = \boldsymbol{d_i\sigma_e^2} \text{---------} (44)$$

Where, $\boldsymbol{d_i}$ is the diagonal element of $\boldsymbol{C^{22}}$. The square root of PEV gave the standard error of predictions.

The accuracies of predicted BVs were obtained as the correlation between predicted and true BVs ($\boldsymbol{r_{\hat{u},u}}$). The accuracy as a function of $\boldsymbol{PEV}$ and $\boldsymbol{\sigma_a^2}$ **,** was used to obtain the accuracies of predicted BVs, which can be expressed as follows:

$$\boldsymbol{r_{\hat{u},u}} = \sqrt{\mathbf{1} - \frac{\boldsymbol{PEV}}{\boldsymbol{\sigma_a^2}}} \text{ ---------} \quad (45)$$

The breeding values for individual animals, the standard error of predictions, and the accuracy of predicted BVs for different traits were obtained by using run "RUNOP –BLUP" in Wombat (Meyer, 2007).

### 3.7.3 Multivariate Best Linear Unbiased Predictor (MBLUP)

The BV prediction via multivariate approach was implemented under the simplest situation of the equal design matrix, where all the traits in the model are affected by the same effects and without missing values where each of the $\boldsymbol{k}$ traits are measured once in each of $\boldsymbol{n}$ individuals. The LMM was constructed for the multivariate data by organising $\boldsymbol{k}$ random vectors for each trait of dimensions $\boldsymbol{n \times 1}$ into a $\boldsymbol{(nk) \times 1}$ dimensional column vector by stacking $\boldsymbol{k}$ univariate random vectors. The following illustrates the compact representation of the multivariate animal model:

$$\boldsymbol{y_j} = \boldsymbol{X_j b_j} + \boldsymbol{Z_j u_j} + \boldsymbol{e_j} \text{ ---------} \quad (46)$$

Where, $\boldsymbol{y_j}$ corresponds to the observation of trait $\boldsymbol{j}$ in the $\boldsymbol{i^{th}}$ individual, $\boldsymbol{b_j}$ represents $\boldsymbol{q_j \times 1}$ fixed effects associated with trait $\boldsymbol{j}$, $\boldsymbol{u_j}$ is $\boldsymbol{p_j \times 1}$ random vector of animal effects of trait $\boldsymbol{j}$ and $\boldsymbol{e_j}$ is the residual error associated with trait $\boldsymbol{j}$ in the $\boldsymbol{i^{th}}$ individual. $\boldsymbol{X_j}$ is $\boldsymbol{n \times q_j}$ design matrix associating $\boldsymbol{q_j}^{\boldsymbol{th}}$ fixed effect with trait $\boldsymbol{j}$ and $\boldsymbol{Z_j}$ is $\boldsymbol{n \times p_j}$ design matrix connecting $\boldsymbol{p_j}^{\boldsymbol{th}}$ random animal effect with trait $\boldsymbol{j}$.

For $\boldsymbol{k}$ traits each measured once in each of $\boldsymbol{n}$ individuals, the above multivariate animal model can be represented with appropriate dimensions of corresponding model components as:

$$\boldsymbol{y_{nk\times 1}} = \boldsymbol{X_{nk\times qk}}\boldsymbol{b_{qk\times 1}} + \boldsymbol{Z_{nk\times nk}}\boldsymbol{u_{nk\times 1}} + \boldsymbol{e_{nk\times 1}} \text{ --------- } (47)$$

$$\boldsymbol{u_{nk\times 1}} \sim \boldsymbol{MVN}(\mathbf{0}, \boldsymbol{G_{nk\times nk}}) \text{ --------- } (48)$$

$$\boldsymbol{e_{nk\times 1}} \sim \boldsymbol{MVN}(\mathbf{0}, \boldsymbol{R_{nk\times nk}}) \text{ --------- } (49)$$

The following assumptions were made about the form of $\boldsymbol{G}$ and $\boldsymbol{R}$ matrices:

$$\boldsymbol{G_{nk\times nk}} = \boldsymbol{C_{k\times k}} \otimes \boldsymbol{A_{n\times n}} \text{ --------- } (50)$$

Where, $\boldsymbol{G}$ is $\boldsymbol{nk} \times \boldsymbol{nk}$ variance-covariance matrix of $\boldsymbol{u_{nk\times 1}}$, $\boldsymbol{C}$ is the $\boldsymbol{k} \times \boldsymbol{k}$ matrix of additive genetic covariances between different traits within an individual, $\boldsymbol{A}$ is $\boldsymbol{n} \times \boldsymbol{n}$ matrix of additive genetic covariances between traits among individuals.

$$\boldsymbol{R_{nk\times nk}} = \boldsymbol{E_{k\times k}} \otimes \boldsymbol{I_{n\times n}} \text{ --------- } (51)$$

Where, $\boldsymbol{R}$ is the $\boldsymbol{nk} \times \boldsymbol{nk}$ covariance matrix of error vector $\boldsymbol{e}$, $\boldsymbol{E}$ is the $\boldsymbol{k} \times \boldsymbol{k}$ is the residual covariances between different characters measured in the same individual and $\boldsymbol{I}$ is the $\boldsymbol{n} \times \boldsymbol{n}$ identity matrix.

The operator $\otimes$ represents the Kronecker product between the matrices.

The MME was used to solve the multivariate animal models to obtain the predictions of BVs for multiple traits. The modified MME for the multivariate animal model is represented as follows:

$$\begin{bmatrix} \boldsymbol{X_j'}(\boldsymbol{E^{-1}} \otimes \mathbf{I})\boldsymbol{X_j} & \boldsymbol{X_j'}(\boldsymbol{E^{-1}} \otimes \mathbf{I}) \\ (\boldsymbol{E^{-1}} \otimes \mathbf{I})\boldsymbol{X_j} & (\boldsymbol{E^{-1}} \otimes \mathbf{I})\boldsymbol{X_j} + (\boldsymbol{C^{-1}} \otimes \boldsymbol{A^{-1}}) \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{b}}_j \\ \widehat{\boldsymbol{u}}_j \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_j'}(\boldsymbol{E^{-1}} \otimes \mathbf{I})\boldsymbol{y_j} \\ (\boldsymbol{E^{-1}} \otimes \mathbf{I})\boldsymbol{y_j} \end{bmatrix} \text{ --------- } (52)$$

### 3.7.4 Accuracy of multivariate breeding values

The accuracy of predicted breeding values from multivariate models was obtained by modifying the equation described under the accuracy of univariate predictions, which can be represented as follows:

$$r_{\hat{u},u} = \sqrt{1 - \frac{PEV_{ij}}{\sigma^2_{a(jj)}}} \text{ ---------} (53)$$

Where, $PEV_{ij}$ is the diagonal element of the coefficient matrix pertaining to animal $i$ and trait $j$ and $\sigma^2_{a(jj)}$ is the additive genetic variance of trait $j$.

The multivariate animal model was implemented in software Wombat (Meyer, 2007). The $C_{k\times k}$ and $E_{k\times k}$ matrices for the full multivariate (6 traits) were obtained through respective bivariate analysis of traits.

### 3.8 Longitudinal data analysis

The longitudinal data recorded on the body weight was analyzed using Multiple-Trait Model and Random Regression Models.

### 3.8.1 Multiple-trait model

The following multiple-trait model was used, following Schaeffer (1984) for analysis of body weights recorded at five different time intervals

$$y = X\beta + Za + Zp_E + e \text{ ---------} (54)$$

Where, $y_i = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} = \begin{pmatrix} Observation\ vector\ for\ BW1 \\ Observation\ vector\ for\ BW2 \\ \vdots \\ \vdots \\ Observation\ vector\ for\ BW5 \end{pmatrix}$ --------- (55)

$$\boldsymbol{X\beta} = \begin{pmatrix} \boldsymbol{X_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X_2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{X_5} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \\ \vdots \\ \boldsymbol{\beta_5} \end{pmatrix} \text{--------- (56)}$$

$\boldsymbol{\beta}_i$ is the fixed effects for the trait $\boldsymbol{i}$ and $\boldsymbol{X_i}$ is the design matrix relating the vector $\boldsymbol{\beta_i}$ and $\boldsymbol{y_i}$. The fixed effects $\boldsymbol{\beta_i}$ were same for all the traits, but the size of the $\boldsymbol{X_i}$ was different for each $\boldsymbol{y_i}$ depending on the number of observations per trait.

$$\boldsymbol{Z(a + p_E)} = \begin{pmatrix} \boldsymbol{Z_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z_2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{Z_5} \end{pmatrix} \begin{pmatrix} \boldsymbol{a_1} & + & \boldsymbol{p_{e1}} \\ \boldsymbol{a_2} & + & \boldsymbol{p_{e2}} \\ \vdots & \vdots & \vdots \\ \boldsymbol{a_5} & + & \boldsymbol{p_{e5}} \end{pmatrix} \text{--------- (57)}$$

Where, $\boldsymbol{a_i}$ is the vector of random animal genetic effect for trait $\boldsymbol{i}$, $\boldsymbol{p_i}$ is the corresponding vector of random permanent environmental effect for trait $\boldsymbol{i}$, $\boldsymbol{Z_i}$ is the design matrix of zeros and ones. If animals are measured only once for particular trait then $\boldsymbol{p_i}$ is confounded with error vector $\boldsymbol{e_i}$.

$$\boldsymbol{E(y) = X\beta} \text{--------- (58)}$$

$$\boldsymbol{E(a) = 0, E(p_e) = 0, E(e) = 0} \text{--------- (59)}$$

$$\boldsymbol{var}\begin{pmatrix} \boldsymbol{a_i} \\ \boldsymbol{p_{ei}} \\ \boldsymbol{e_i} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A \otimes G} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I \otimes P} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I \otimes R} \end{pmatrix} \text{--------- (60)}$$

Where, $\boldsymbol{A}$ is the numerator relationship matrix, $\boldsymbol{G}$ is the matrix of genetic variances and covariances among traits, $\boldsymbol{P}$ is the matrix of variances and covariances of permanent environmental effects among traits, $\boldsymbol{R}$ is the matrix of error variances and covariances among traits, $\boldsymbol{I}$ is the identity matrix and $\otimes$ denotes the direct product of two matrices.

If, $\boldsymbol{G} = \begin{pmatrix} \boldsymbol{g_{11}} & \boldsymbol{g_{12}} & \cdots & \boldsymbol{g_{15}} \\ \boldsymbol{g_{12}} & \boldsymbol{g_{22}} & \cdots & \boldsymbol{g_{25}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{g_{15}} & \boldsymbol{g_{25}} & \cdots & \boldsymbol{g_{55}} \end{pmatrix}$, then, $\boldsymbol{var}\begin{pmatrix} \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \vdots \\ \boldsymbol{a_5} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Ag_{11}} & \boldsymbol{Ag_{12}} & \cdots & \boldsymbol{Ag_{15}} \\ \boldsymbol{Ag_{12}} & \boldsymbol{Ag_{22}} & \cdots & \boldsymbol{Ag_{25}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{Ag_{15}} & \boldsymbol{Ag_{25}} & \cdots & \boldsymbol{Ag_{55}} \end{pmatrix}$**----** (61)

If, $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{p_{11}} & \boldsymbol{p_{12}} & \cdots & \boldsymbol{p_{15}} \\ \boldsymbol{p_{12}} & \boldsymbol{p_{22}} & \cdots & \boldsymbol{p_{25}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{p_{15}} & \boldsymbol{p_{25}} & \cdots & \boldsymbol{p_{55}} \end{pmatrix}$, then, $\boldsymbol{var}\begin{pmatrix} \boldsymbol{p_{e1}} \\ \boldsymbol{p_{e2}} \\ \vdots \\ \boldsymbol{p_{e5}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Ip_{11}} & \boldsymbol{Ip_{12}} & \cdots & \boldsymbol{Ip_{15}} \\ \boldsymbol{Ip_{12}} & \boldsymbol{Ip_{22}} & \cdots & \boldsymbol{Ip_{25}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{Ip_{15}} & \boldsymbol{Ip_{25}} & \cdots & \boldsymbol{Ip_{55}} \end{pmatrix}$**----** (62)

If, $\boldsymbol{R} = \begin{pmatrix} \boldsymbol{r_{11}} & \boldsymbol{r_{12}} & \cdots & \boldsymbol{r_{15}} \\ \boldsymbol{r_{12}} & \boldsymbol{r_{22}} & \cdots & \boldsymbol{r_{25}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{r_{15}} & \boldsymbol{r_{25}} & \cdots & \boldsymbol{r_{55}} \end{pmatrix}$, then, $\boldsymbol{var}\begin{pmatrix} \boldsymbol{e_1} \\ \boldsymbol{e_2} \\ \vdots \\ \boldsymbol{e_5} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Ir_{11}} & \boldsymbol{Ir_{12}} & \cdots & \boldsymbol{Ir_{15}} \\ \boldsymbol{Ir_{12}} & \boldsymbol{Ir_{22}} & \cdots & \boldsymbol{Ir_{25}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{Ir_{15}} & \boldsymbol{Ir_{25}} & \cdots & \boldsymbol{Ir_{55}} \end{pmatrix}$**----** (63)

### 3.8.2 Random regression analysis

In each batch, the families were produced in an interval of 3 months, which resulted in 10 different age classes, subsequently resulting in 50 different age class of body weight records when sampled across five different time intervals. To model the shape of longitudinal data across 50 ages, the orthogonal Legendre polynomials were used as suggested by Schaeffer (2016). The age data were mapped to a scale between -1.0 to 1.0 in order to convert covariate age into Legendre polynomial. The following scaling formula was used for the mapping.

$$\boldsymbol{q_i} = \boldsymbol{-1 + 2\left(\frac{t_i - t_{min}}{t_{max} - t_{min}}\right)} \text{ --------- } (64)$$

Where $\boldsymbol{q_i}$ is the scaled decimal number corresponding to age $\boldsymbol{t_i}$, $\boldsymbol{t_{min}}$ is the starting point, $\boldsymbol{t_{max}}$ is the end point, $\boldsymbol{t_i}$ is the whole number value of age to be mapped to $\boldsymbol{q_i}$.

### 3.8.2.1 Legendre polynomials

The Legendre polynomials were defined as a function of standardized age $\boldsymbol{q_i}$ follows, as adapted from Schaeffer (2016). The first two Legendre polynomials are defined as:

$$\boldsymbol{P_0(x) = 1}, and\ \boldsymbol{P_1(x) = x} \text{ --------- } (65)$$

Where $\boldsymbol{x}$ is one of the $\boldsymbol{q_i}$.

Then, in general, the $(\boldsymbol{n+1})^{\boldsymbol{th}}$ polynomial can be described by the following recursive formula:

$$\boldsymbol{P_{(n+1)}(x) = \frac{1}{(n+1)}\{(2n+1)xP_n(x) - nP_{n-1}(x)\}} \text{---------} (66)$$

These values are then normalized using:

$$\boldsymbol{\phi_n(x) = \left(\frac{2n+1}{2}\right)^{0.5} P_n(x)} \text{ --------- } (67)$$

Thus the following normalized quadratic Legendre polynomials were obtained:

$$\boldsymbol{n = 0, P_0(x) = 1, \phi_0(x) = 0.7071} \text{---------} (68)$$

$$\boldsymbol{n = 1, P_1(x) = x, \phi_1(x) = 1.2247x} \text{---------} (69)$$

$$\boldsymbol{n = 2, P_2(x) = \frac{3}{2}x^2 - \frac{1}{2}, \phi_2(x) = -0.7906 + 2.3717x^2} \text{---------} (70)$$

Then the $3 \times 3$ matrix of coefficients of normalized quadratic Legendre polynomials is:

$$\boldsymbol{\Lambda_{3\times3}} = \begin{pmatrix} \boldsymbol{0.7071} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1.2247} & \boldsymbol{0} \\ \boldsymbol{-0.7906} & \boldsymbol{0} & \boldsymbol{2.3717} \end{pmatrix} \text{---------} (71)$$

Applying equation $(64)$ on 50 different ages resulted in a matrix $\boldsymbol{M}$ of dimension $50 \times 3$.

$$\boldsymbol{M_{50\times 3}} = \begin{pmatrix} \mathbf{1} & \boldsymbol{q_1} & \boldsymbol{q_1^2} \\ \mathbf{1} & \boldsymbol{q_2} & \boldsymbol{q_2^2} \\ \vdots & \vdots & \vdots \\ \mathbf{1} & \boldsymbol{q_{50}} & \boldsymbol{q_{50}^2} \end{pmatrix} \text{---------} (72)$$

Where the first column of the matrix $\boldsymbol{M}$ is a column of ones representing the intercept of the curve, the second column is the standardized age corresponding to the linear term, while the third column is the standardized age squared for the quadratic term. As the order of the polynomial increases, the number of columns of matrix $\boldsymbol{M}$ also increases.

Then the standardized ages ($\boldsymbol{M_{50\times 3}}$) and coefficients of normalized polynomials ($\boldsymbol{\Lambda_{3\times 3}}$) were combined to produce the matrix $\boldsymbol{\Phi}$ containing the Legendre polynomials evaluated at each of the observed ages.

$$\boldsymbol{\Phi_{50\times 3}} = \boldsymbol{M_{50\times 3}\Lambda_{3\times 3}} \text{---------} (73)$$

### 3.8.2.2 Covariance function

The variances and covariances between 50 different ages were estimated as follows:

$$\boldsymbol{V_{50\times 50}} = \boldsymbol{\Phi_{50\times 3}}\mathbf{K_{3\times 3}}\boldsymbol{\Phi'_{3\times 50}} \text{---------} (74)$$

Where $\boldsymbol{V_{50\times 50}}$ is the matrix of variances and covariances between 50 different ages considered in the analysis, $\boldsymbol{\Phi_{50\times 3}}$ are the Legendre polynomials evaluated at each age, $\mathbf{K_{3\times 3}}$ is the $3 \times 3$ coefficients of covariance function matrix of variances and covariances between the intercept, linear and quadratic terms of the Legendre polynomials. The matrix $\mathbf{K_{3\times 3}}$ was estimated using the REML procedure in software Wombat (Meyer, 2007).

### 3.8.2.3 Random regression model (RRM)

Two different RRMs were used under the assumptions of residual homogeneity of variance (HOM), residual homogeneity of variance. In its most general form, an RRM including family effect and residual heterogeneity of variance in matrix notations is written as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z_1 u} + \boldsymbol{Z_2 p_e} + \boldsymbol{e} \text{ ---------} (75)$$

Where $\boldsymbol{y}$ is the vector of observations, $\boldsymbol{b}$ is the vector of regression coefficients of fixed effects; $\boldsymbol{u}$, and $\boldsymbol{p_e}$ are the vectors of random regression coefficients of additive genetic and permanent environmental effects respectively; $\boldsymbol{X}$, $\boldsymbol{Z_1}$, and $\boldsymbol{Z_2}$ are design matrices relating vectors $\boldsymbol{b}$, $\boldsymbol{u}$, and $\boldsymbol{p_e}$ including Legendre polynomials obtained as functions of age (equation (73)) to the observation vector $\boldsymbol{y}$ and $\boldsymbol{e}$ is the vector of residual errors.

The following assumptions were made:

$$\boldsymbol{E(y)} = \boldsymbol{X\beta} \text{ ---------} (76)$$

$$\boldsymbol{E(u)} = \boldsymbol{E(p_e)} = \boldsymbol{E(e)} = \boldsymbol{0} \text{ ---------} (77)$$

For heterogeneous variance model (HET):

$$\boldsymbol{var(y)} = \boldsymbol{Z_1(A \otimes G)Z_1'} + \boldsymbol{Z_2(I \otimes P)Z_2'} + \boldsymbol{I\sigma_{ek}^2} \text{ ---------} (78)$$

For homogeneous variance model (HOM):

$$\boldsymbol{var(y)} = \boldsymbol{Z_1(A \otimes G)Z_1'} + \boldsymbol{Z_2(I \otimes P)Z_2'} + \boldsymbol{I\sigma_e^2} \text{ ---------} (79)$$

Where $\boldsymbol{A}$ is the additive genetic relationship matrix; $\boldsymbol{G}$, and $\boldsymbol{P}$ and are $3 \times 3$ matrix of coefficients of covariance functions which are the variances and covariances corresponding to the intercept, linear and quadratic terms of random regression coefficients associated with additive genetic, and permanent environmental effects

respectively; $\boldsymbol{\sigma^2_{ek}}$ is the heterogeneous residual variances for $\boldsymbol{k}$ classes (where $\boldsymbol{k = 50}$); $\boldsymbol{\sigma^2_e}$ is the homogeneous residual variance; $\boldsymbol{I}$ is the identity matrix corresponding to each effect and ⊗ is the Kronecker product.

The elements of matrix $\boldsymbol{G}$, and $\boldsymbol{P}$ as well as $\boldsymbol{\sigma^2_{ek}}$ and $\boldsymbol{\sigma^2_e}$ were estimated by using the software Wombat (Meyer, 2007) using the REML procedure. The matrices $\boldsymbol{G}$, and $\boldsymbol{P}$ contains the respective coefficients of covariance functions associated with the particular random effects. Applying these estimators to equation $(74)$ by replacing $\mathbf{K_{3\times3}}$ by $\boldsymbol{G}$, and $\boldsymbol{P}$ gave the full variance covariance matrix of dimension $50 \times 50$ of respective random effects for the age range considered in the analysis.

The heritability from the respective models at age $\boldsymbol{i}$ was obtained as follows:

$$\boldsymbol{h^2_{HOM}} = \frac{\boldsymbol{Z_{1i}GZ'_{1i}}}{\boldsymbol{Z_{1i}GZ'_{1i}+Z_{2i}PZ'_{2i}+\sigma^2_e}} \text{ --------- } (80)$$

$$\boldsymbol{h^2_{HET}} = \frac{\boldsymbol{Z_{1i}GZ'_{1i}}}{\boldsymbol{Z_{1i}GZ'_{1i}+Z_{2i}PZ'_{2i}+\sigma^2_{ek}}} \text{ --------- } (81)$$

The ratios of other variance components were estimated by replacing the numerator with appropriate random effects in equations $(80)$ and $(81)$**.**

The genetic correlation between any two ages $\boldsymbol{i}$ and $\boldsymbol{j}$ was estimated in the following way:

$$\boldsymbol{r_{ij}} = \frac{\boldsymbol{Z_{1i}GZ'_{1j}}}{\sqrt{\boldsymbol{Z_{1i}GZ'_{1j}}}\sqrt{\boldsymbol{Z_{1i}GZ'_{1j}}}} \text{ --------- } (82)$$

The correlations between permanent environmental and common environmental effects were estimated by replacing the equation $(82)$ with appropriate random effects.

### 3.8.2.4 Eigenfunctions

The shape of eigenfunctions and the magnitude of the associated eigenvalues provide information about the possible change of the growth curve (Van der Werf, 2002). The eigenfunctions were estimated following Van der Werf (2002) by first decomposing the genetic covariance function matrix into its corresponding eigenvalues and eigenvectors as follows:

$$\boldsymbol{K} = \boldsymbol{EDE'} \text{ ---------} \quad (83)$$

Where $\boldsymbol{K}$ is the genetic covariance function estimated after fitting the model represented in $(75)$**;** $\boldsymbol{E}$ is the eigenvectors of $\boldsymbol{K}$, and $\boldsymbol{D}$ is the eigenvalues of $\boldsymbol{K}$.

The eigenfunctions of $\boldsymbol{K}$ is then:

$$\boldsymbol{Q} = \boldsymbol{\Phi E} \text{ ---------} \quad (84)$$

Where the $\boldsymbol{k}$ columns of $\boldsymbol{Q}$ represent the $\boldsymbol{k}$ eigenfunctions with associated eigenvalues $\boldsymbol{d_i}$, the matrix $\boldsymbol{\Phi}$ containing the Legendre polynomials evaluated at each of the observed ages from the equation $(73)$ and $\boldsymbol{E}$ is the eigenvectors of $\boldsymbol{K}$. Associated with each eigenfunction is a canonical variable $\boldsymbol{z_i}$ with variance $\boldsymbol{d_i}$ and single trait selection on each $\boldsymbol{z_i}$ gives a response along the trajectory proportional to its eigenfunction (Van der Werf, 2002).

# 4. RESULTS

## 4.1 Population structure

The population structure with different factors and the number of families within each factor is illustrated in Table 1. The number of fish stocked per pond and the number of fish at the time of harvest are also listed in Table 4.1. The data consisted of records from 78 full-sib families, 39 families each from batch-1 and batch-2. Batch-1 fish were produced from three stocks, namely Andhra Pradesh (AP), Assam and West Bengal. From AP stock, 23 families were produced in batch-1, whereas 16 families were produced from Assam and West Bengal stocks in batch-1. Due to fewer families from Assam and West Bengal stocks, and to circumvent any possible issues in the estimation, they were combined and coded as Non-Andhra Pradesh stock (NAP stock). All the families from batch-2 were exclusively from AP-stock. The communal rearing for batch-1 was carried out under both monoculture and polyculture system, whereas batch-2 communal rearing was restricted to polyculture only. Families of batch-1 were distributed among 5 ponds and the batch-2 families were distributed across 6 ponds, different ponds were used for culture purpose in both the batches. A total of 2328 PIT-tagged fish were stocked for communal rearing, and 1413 were recovered at the harvest a reduction of 39 percent. The number of fish stocked per pond and the number of fish harvested from each pond are given in Table 4.1.

## 4.2 Least squares means

Least squares means and their standard errors, and coefficient of variation (CV %), for different factors for various traits at stocking and harvest are listed in Table 2A. On average, the body weight of fish increased by 108 g from stocking to harvest. At the time of harvest, mean body weight from different ponds ranged from 101 to 156 g (Table 2B). Mean squares from the analysis of variance for different traits at stocking and harvest are presented in Tables 3 and 4, respectively.

### 4.3 Effect of non-genetic factors

The population structure with different non-genetic factors and the number of families per factor are presented in Table 1. For the traits at stocking, effects such as batch, stock, and sex were considered as non-genetic effects. The effect of the batch was significant for BW0 and TL0, whereas the stock effect was non-significant. The average BW0 of males was significantly higher than the average BW0 of females. Different non-genetic factors considered for harvest traits were initial body weight (covariate), stock, batch, culture-type, pond, and sex, and all these factors were considered as fixed effects in model-1. The type-III Mean sum of squares and coefficient of determination ($R^2$%) obtained after fitting model-1, along with the level of significance for traits at stocking and harvest, are presented in Tables 3 and 4, respectively. There was no significant difference in the mean performance of harvest traits between fish cultured under monoculture and polyculture systems (Table 3). The batch effect was not significant for TL at harvest, and the sex effect was not significant for BD at harvest. The body weight at stocking (BW0) had a significant effect on all the traits at harvest.

There were two levels for the stock, AP stock and Non-AP stock. On average, the fish from Non-AP stock grew 13.5 g heavier and 0.5 cm longer than the fish from AP stock (Table 2A). The average harvest BW and TL of fish from batch-1 were higher than that for the fish from batch-2 by 16.5 g and 1.8 cm, respectively. However, ADG and BD were higher in fish from batch-2. In magur, the males tend to grow heavier than the females, and in the present study, on average, males grew 27.31 g heavier than females at harvest (Table 3).

Across the two batches, fish were reared in eleven different earthen ponds. The fish from batch-1 were reared in five different ponds, and batch-2 fish were reared in another set of six different ponds. The pond had a profound effect on the traits at harvest. Within batch-1, LS means for harvest BW among five ponds ranged from 135.84 to 151.5 g, whereas for the six ponds within batch-2, it ranged from 101.4 to 156.7 g (Table 2B).

Table 1. Population structure with different factors and number of families per factor. (Batch-1 – 2014 year class, Batch-2 – 2015 year class; AP – Andhra Pradesh Stock, NAP – Non-Andhra Pradesh Stock); N – Number of fish per pond. The bold letters indicate different levels of factors, and the numbers indicate the number of families within each level of the factor. The numbers in the last two rows indicate the number of fish per pond at the time of stocking and harvest, respectively.

| **Factors and No. of Families per Factor** | **Batch-1** – 39 | | | | | **Batch-2** - 39 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **AP Stock** – 23<br>**NAP Stock** – 16 | | | | | **AP Stock** - 39 | | | | | |
| | **Monoculture** - 39 | | | **Polyculture** - 39 | | **Polyculture** - 39 | | | | | |
| | **P 1** | **P 3** | **P 4** | **P 2** | **P 5** | **P 6** | **P 7** | **P 8** | **P 9** | **P 10** | **P 11** |
| | 39 | 39 | 39 | 33 | 29 | 13 | 13 | 12 | 13 | 13 | 13 |
| **N (Stocking)** | 310 | 309 | 309 | 200 | 230 | 183 | 193 | 192 | 126 | 154 | 122 |
| **N (Harvest)** | 232 | 232 | 236 | 101 | 148 | 101 | 79 | 76 | 53 | 71 | 84 |

Table 2A: Number of observations ($N_S$ – for stocking and $N_H$ – for harvest), overall, stock, batch, and sex-wise least squares means and their standard errors for the traits (BW0 and TL0) at stocking and harvest (BW, TL, BD, HW, and ADG)

| Effects | $N_S$ | BW0 (g) | TL0 (cm) | $N_H$ | BW (g) | TL (cm) | BD (cm) | HW (cm) | K ($g/cm^3$) | ADG (g/day) |
|---|---|---|---|---|---|---|---|---|---|---|
| **Overall** | 2328 | 26.11±0.20 | 14.12±0.05 | 1413 | 134.92±0.99 | 24.49±0.07 | 2.92±0.02 | 3.87±0.02 | 0.91±0.01 | 0.28±0.01 |
| **CV (%)** | | 36.90 | 17.10 | 2328 | 27.48 | 10.67 | 15.28 | 10.57 | 18.55 | 28.39 |
| **Stock** | | | | | | | | | | |
| **AP** | 1763 | 25.7±0.28[a] | 14.16±0.06[a] | 983 | 127.15±1.01[a] | 23.97±0.07[a] | 2.88±0.02[a] | 3.82±0.02[a] | 0.92±0.01[a] | 0.28±0.01[a] |
| **NAP** | 565 | 26.8±0.56[a] | 14.16±0.11[a] | 430 | 140.61±1.78[b] | 24.43±0.12[b] | 3.05±0.03[b] | 3.92±0.03[b] | 0.95±0.01[b] | 0.3±0.01[b] |
| **Batch** | | | | | | | | | | |
| **1** | 1358 | 41.78±0.32[a] | 17.76±0.07[a] | 949 | 142.14±1.06[a] | 25.09±0.08[a] | 2.93±0.02[a] | 3.89±0.02[a] | 0.89±0.01[a] | 0.27±0.01[a] |
| **2** | 970 | 10.72±0.49[b] | 10.56±0.10[b] | 464 | 125.62±1.76[b] | 23.31±0.12[b] | 2.99±0.03[a] | 3.85±0.03[a] | 0.98±0.01[b] | 0.3±0.01[b] |
| **Sex** | | | | | | | | | | |
| **Male** | 654 | 29.3±0.42[a] | 14.7±0.08[a] | 654 | 147.54±1.30[a] | 25.2±0.09[a] | 3.03±0.02[a] | 4.04±0.02[a] | 0.92±0.01[a] | 0.32±0.01[a] |
| **Female** | 759 | 23.7±0.44[b] | 13.8±0.08[b] | 759 | 120.23±1.36[b] | 23.19±0.10[b] | 2.89±0.02[b] | 3.7±0.02[b] | 0.96±0.01[b] | 0.26±0.01[b] |
| **NA** | 915 | 25.2±0.48 [b] | 13.7±0.09 [b] | - | - | - | - | - | | - |

Means bearing the same superscript within the class effect are not significantly different ($p < 0.01$) from one another.

Table 2B: Pond wise (within a batch) least squares means from batch-1and2 for different traits at harvest; N is the number of observations

| Batch | Pond | N (Stocking) | N (Harvest) | BW (g) | TL (cm) | BD (cm) | HW (cm) | K ($g/cm^3$) | ADG (g/day) |
|---|---|---|---|---|---|---|---|---|---|
| **1** | **1** | 310 | 232 | 135.84±2.01 | 24.49±0.14 | 2.65±0.03 | 3.76±0.03 | 0.92±0.02 | 0.26±0.01 |
| **1** | **2** | 200 | 101 | 137.48±3.04 | 24.74±0.21 | 2.65±0.04 | 3.86±0.04 | 0.89±0.02 | 0.26±0.01 |
| **1** | **3** | 309 | 232 | 151.56±2.01 | 25.45±0.14 | 2.85±0.03 | 3.96±0.03 | 0.91±0.02 | 0.30±0.01 |
| **1** | **4** | 309 | 236 | 148.57±1.99 | 26.12±0.14 | 3.30±0.03 | 4.00±0.03 | 0.83±0.02 | 0.29±0.01 |
| **1** | **5** | 230 | 148 | 137.27±2.51 | 24.67±0.17 | 3.20±0.04 | 3.88±0.03 | 0.92±0.02 | 0.26±0.01 |
| **2** | **6** | 183 | 101 | 101.37±3.20 | 21.45±0.22 | 2.71±0.04 | 3.60±0.04 | 1.02±0.02 | 0.25±0.01 |
| **2** | **7** | 193 | 79 | 126.91±3.57 | 23.45±0.24 | 3.02±0.05 | 3.89±0.05 | 0.97±0.02 | 0.31±0.01 |
| **2** | **8** | 192 | 76 | 128.63±3.64 | 23.37±0.25 | 3.01±0.05 | 3.82±0.05 | 0.99±0.02 | 0.32±0.01 |
| **2** | **9** | 126 | 53 | 156.74±4.30 | 25.22±0.29 | 3.21±0.06 | 4.14±0.05 | 0.97±0.03 | 0.39±0.01 |
| **2** | **10** | 154 | 71 | 117.03±3.75 | 22.87±0.26 | 3.04±0.05 | 3.81±0.05 | 0.97±0.02 | 0.27±0.01 |
| **2** | **11** | 122 | 84 | 123.05±3.47 | 23.48±0.24 | 2.96±0.05 | 3.85±0.05 | 0.94±0.02 | 0.29±0.01 |

Table 3: Mean sum of squares for model parameters and $R^2$ values of model for traits at stocking

| MSS at stocking | | | | |
|---|---|---|---|---|
| **Source** | **DF** | **BW0** | **TL0** | **K0** |
| **Stock** | 1 | 291.92 | 1.59 | 0.04 |
| **Batch** | 1 | 369846.60** | 19356.27** | 1.20** |
| **Sex** | 2 | 3888.90** | 111.33** | 0.21** |
| **Family** | 75 | 1524.03** | 75.96** | 7.10** |
| **Error** | 2248 | 77.50 | 2.60 | 0.02 |
| **$R^2$ (%)** | | 80.00 | 86.00 | 22.00 |

**** (p-value <0.01)

Table 4: Mean sum of squares for model parameters and $R^2$ values of model for traits at harvest

| MSS at harvest | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Source** | **DF** | **BW** | **TL** | **BD** | **HW** | **K** | **ADG** |
| **BW0** | 1 | 244716.34** | 822.23** | 18.01** | 20.85** | 0.01 | 0.29** |
| **Stock** | 1 | 21641.19** | 23.51** | 3.78** | 0.86** | 0.13 | 0.15** |
| **Batch** | 1 | 12137.07** | 24.39** | 1.09** | 2.16** | 0.07 | 0.08** |
| **Culture type** | 1 | 347.82 | 2.94 | 0.09 | 0.12 | 0.01 | 0.01 |
| **Pond** | 9 | 10832.34** | 60.04** | 8.09** | 1.37** | 0.18** | 0.08** |
| **Sex** | 1 | 69676.30** | 505.30** | 0.35 | 15.25** | 0.56** | 0.45** |
| **Family** | 75 | 2785.61** | 11.18** | 0.31** | 0.28** | 0.05** | 0.02** |
| **Error** | 1323 | 453.30 | 2.60 | 0.09 | 0.08 | 0.02 | 0.004 |
| **$R^2$ (%)** | | 69.00 | 65.00 | 55.00 | 57.00 | 18.00 | 51.00 |

**** (p-value <0.01)

## 4.4 Heritability estimates

Heritability for the traits at harvest was estimated by adopting both models 1 and 2; the additive genetic variance and residual variance of harvest traits estimated from both models 1 and 2 were similar. Estimates of heritability along with their standard error for traits at stocking and harvest are presented in Table 5. The heritability estimated for BW, ADG (0.42 – 0.44) at harvest was high, whereas a moderate heritability was obtained for TL, BD, and HW (0.22 to 0.32) and a very low heritability for K (0.07). The heritability of the traits at stocking was very high, 0.74 for BW0 and 0.97 for TL, while moderate (0.35) for K0.

## 4.5 Genetic and phenotypic correlations

The estimates of genetic and phenotypic correlations with approximate sampling variance obtained from model-1, among all traits at stocking and harvest, are presented in Table 6**.** The genetic correlations among traits at harvest were positive and high. For traits at harvest, the highest genetic correlation was found between BW and ADG and the lowest between TL and HW. The genetic correlation between BW0 and TL0 was positive and high. All phenotypic correlations between traits at harvest were positive and ranged from moderate (0.30 - 0.60) to high (>0.60). The highest phenotypic correlation was observed between BW and TL, whereas the lowest was obtained between TL and BD.

The estimated genetic and phenotypic correlations between traits at stocking and harvest are presented in Table 7. The genetic correlation between body weight at stocking with body weight and total length at harvest was high and positive. Total length at stocking also had a positive but moderate genetic correlation with body weight and length at harvest. Likewise, there was a positive and moderate phenotypic correlation between the body weight and total length at stocking to body weight and total length at harvest.

Table 5: Estimates of heritability and their standard errors at stocking (BW0: Body weight, TL0: Total length) and at harvest (BW: Body weight, TL: Total Length, BD: Body depth, HW: Head width, ADG: Average daily gain)

| **Traits at stocking** | | **Traits at harvest** | | |
|---|---|---|---|---|
| **Traits** | **Model 1** | **Traits** | **Model 1** | **Model 2** |
| **BW0** | 0.74±0.08 | **BW** | 0.44±0.07 | 0.34±0.07 |
| | | **TL** | 0.32±0.06 | 0.26±0.06 |
| **TL0** | 0.97±0.08 | **BD** | 0.22±0.04 | 0.14±0.04 |
| | | **HW** | 0.27±0.05 | 0.23±0.05 |
| **K0** | 0.35±0.05 | **K** | 0.07±0.02 | 0.06±0.02 |
| | | **ADG** | 0.42±0.07 | 0.32±0.07 |

Table 6: Genetic (above diagonal) and phenotypic (below diagonal) correlations between traits at stocking (BW0: Body weight, TL0: Total length, K0: Condition factor) and among traits at harvest (BW: Body weight, TL: Total Length, BD: Body depth, HW: Head width, K: Condition factor ADG: Average daily gain)

| **Traits at tagging** | | **BW0** | **TL0** | **K0** | | | |
|---|---|---|---|---|---|---|---|
| | **BW0** | | 0.92±0.01 | 0.17±0.11 | | | |
| | **TL0** | 0.84±0.01 | | -0.33±0.1 | | | |
| | **K0** | 0.02±0.03 | 0.41±0.02 | | | | |
| **Traits at harvest** | | **BW** | **TL** | **BD** | **HW** | **K** | **ADG** |
| | **BW** | | 0.95±0.03 | 0.79±0.09 | 0.81±0.06 | 0.18±0.2 | 0.79±0.14 |
| | **TL** | 0.78±0.02 | | 0.66±0.12 | 0.75±0.08 | 0.09±0.21 | 0.73±0.16 |
| | **BD** | 0.55±0.03 | 0.46±0.03 | | 0.68±0.12 | 0.2±0.25 | 0.64±0.18 |
| | **HW** | 0.68±0.02 | 0.62±0.02 | 0.49±0.03 | | 0.16±0.21 | 0.65±0.15 |
| | **K** | 0.1±0.05 | 0.38±0.04 | 0.03±0.05 | 0.07±0.04 | | 0.17±0.34 |
| | **ADG** | 0.82±0.04 | 0.63±0.04 | 0.44±0.04 | 0.56±0.04 | 0.12±0.07 | |

Table 7: Genetic and phenotypic correlations between traits at stocking (BW0: Body weight, TL0: Total length) and harvest (BW: Body weight, TL: Total Length)

| | **Genetic correlations** | | **Phenotypic correlations** | |
|---|---|---|---|---|
| | **BW** | **TL** | **BW** | **TL** |
| **BW0** | 0.76±0.06 | 0.73±0.07 | 0.61±0.03 | 0.48±0.04 |
| **TL0** | 0.57±0.09 | 0.57±0.09 | 0.53±0.03 | 0.47±0.04 |

### 4.6 Variance components and heritability estimates using different methods

The additive and residual variance and heritability for harvest body weight were estimated by seven different methods like ANOVA, REML, Bayesian MCMC, resampling methods such as Non-parametric bootstrap, Parametric bootstrap, Asymptotic sampling, and jackknife method. The results obtained from different methods are presented below.

#### 4.6.1 Analysis of Variance (ANOVA) method

The least squares estimates of variance components were estimated with a full-sib model. The estimates of family variance were obtained by equating mean squares to their expected values. The additive genetic variance (268.17), residual variance (453.26), and heritability (0.45 ± 0.08) with standard error are given in Table 9. The additive variance was obtained as twice the family variance (134.08). The phenotypic variance (587.37) was obtained by summing family variance and residual variance.

#### 4.6.2 Residual Maximum Likelihood (REML) method

The results from the REML analysis of variance components and heritability are presented in Table 9. The additive genetic variance (255.2) and the residual variance (325.2) for the body weight were obtained by REML using the average information algorithm, and the corresponding sampling variances were obtained from the inverse of the average information matrix. A heritability of 0.44±0.07, was obtained as the function of variance components, and its sampling distribution was obtained by approximating the function by first-order Taylor series expansion. The corresponding approximate 95 % confidence interval for the variance components and heritability are given in Table 9.

### 4.6.3 Bayesian method

#### 4.6.3.1 Monte Carlo Markov Chain (MCMC) diagnostics

##### 4.6.3.1.1 Convergence and autocorrelation

The time series trace of additive genetic variance, residual variance, and heritability is presented in Fig 1A-C., depicting the plot of variables generated versus the number of iterations. The trace plot shows the fluctuation in sampled values over the iterations (50,000 to 10,00,000 after dropping the first 50,000 iterations known as burn-in). The trace plot depicted a well-mixing chain that moves through the entire parameter space without confining to any particular region. Autocorrelation between additive genetic variances and residual variances, at five different intervals of iterations, is given in Table 8. In Table 8 the Lag 100 represents autocorrelation between the values of the parameters at every 100 iterations. The autocorrelation was found to be 0.07 for additive genetic variance and 0.06 for the residual variance. Hence, it was decided to keep only one iteration value at every 100 iterations, which means a thinning interval of 100. The thinning reduced the sample size and a total of 9500 values were sampled from the chain. Even though the chain length was 10,00,000, the first 50,000 values were dropped for convergence reasons giving rise to a length of 9,50,000, from which sampling was performed at every 100 iterations, which resulted in the final sample size of 9500. The effective sample size with zero correlation was 8129 for additive genetic variance and 8517 for the residual variance.

##### 4.6.3.1.2 High-density regions (HDRs)

The posterior probability distribution of additive genetic variance, residual variance, and heritabilities is presented in Fig 2A-C. The additive genetic variance ranged from 119.7 to 678.3 (Fig 2A), with the probability of 95% that the parameter values were enclosed within the range of 174.5 to 400.3 (Table 9). The residual variance ranged from 201 to 527.9 (Fig 2B), with the 95% HDR ranging from 321.2 to 466.9 (Table 9). The posterior probability means of additive genetic variance and residual variance were 279.4 and 395.4, respectively, and are presented in Table 9.

Heritability estimated from the posterior distribution of variance components ranged from 0.20 to 0.77 and both the mean and median of heritability was 0.41 and mode was 0.43, with a standard error of 0.07 for the mean heritability (Table 9). 95% HDR for heritability ranged between 0.27 to 0.54 which indicates that there is a 95 % probability for the true value of heritability to lie between 0.27 and 0.54.

Table 8: Autocorrelation values for different intervals for the MCMC chain

| **Intervals** | **Additive Variance** | **Residual Variance** |
|---|---|---|
| **Lag 0** | 1.000 | 1.000 |
| **Lag 100** | 0.078 | 0.054 |
| **Lag 500** | 0.007 | 0.007 |
| **Lag 1000** | 0.010 | 0.011 |
| **Lag 5000** | -0.022 | -0.005 |

Table 9: Estimates of variance components, heritabilities with their sampling distributions and confidence intervals

| Method | Functions | Estimate | SE | L - 95% CI | U - 95% CI |
|---|---|---|---|---|---|
| **ANOVA** | Family | 134.08 | | | |
| | Additive | 268.17 | | | |
| | Residual | 453.26 | | | |
| | Heritability | 0.45 | 0.08 | | |
| **REML** | Additive | 255.20 | 51.68 | 153.90 | 356.50 |
| | Residual | 325.20 | 31.90 | 262.70 | 387.70 |
| | Heritability | 0.44 | 0.07 | 0.30 | 0.58 |
| **Bayesian MCMC** | Additive | 279.40(Mean)<br>273.00(Median)<br>268.30(Mode) | 60.30 | 174.50 | 400.30 |
| | Residual | 395.40(Mean)<br>397.40(Median)<br>404.80(Mode) | 37.40 | 321.20 | 467.00 |
| | Heritability | 0.41(Mean)<br>0.41(Median)<br>0.43(Mode) | 0.07 | 0.28 | 0.55 |

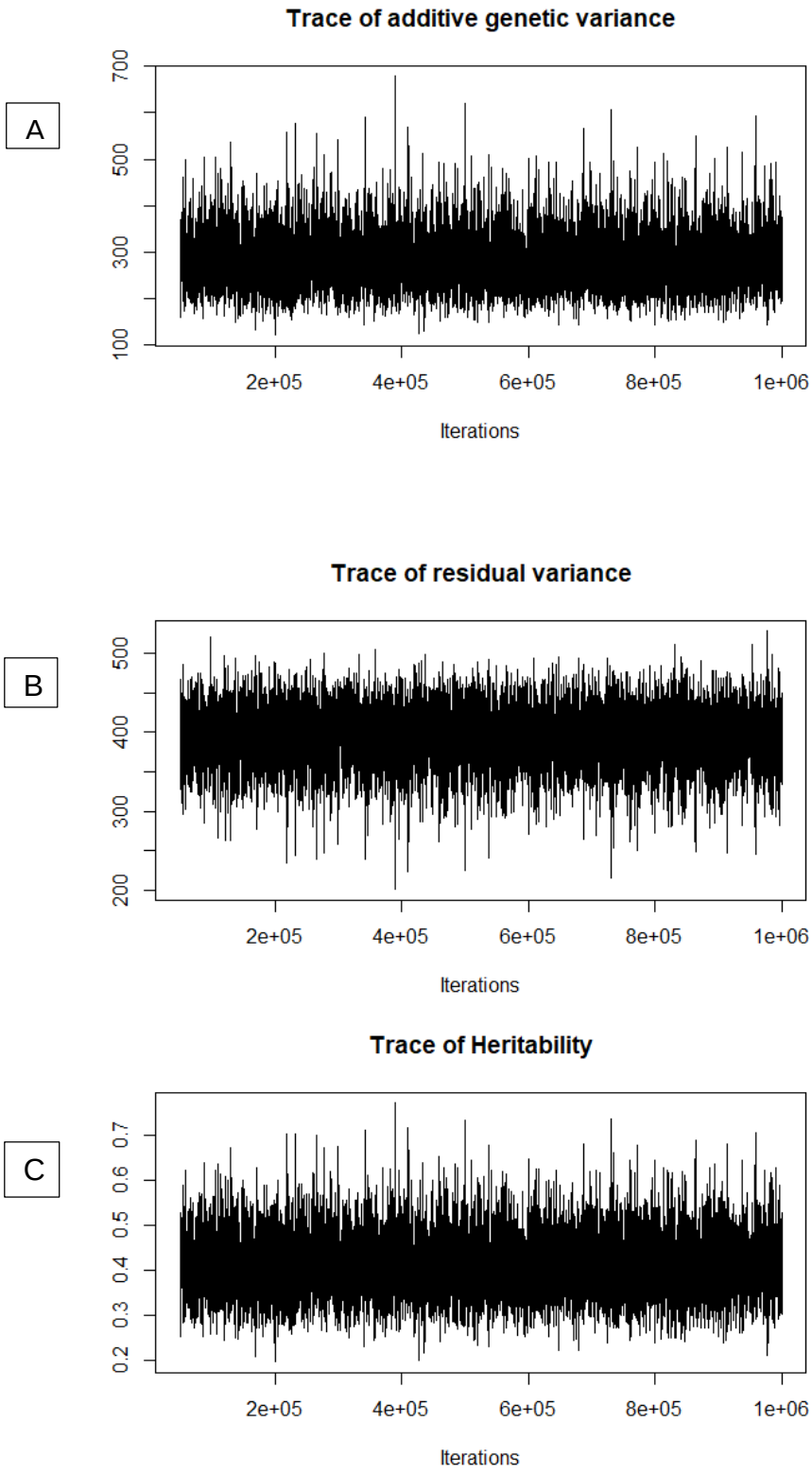



Figure 1: Evolution of time series trace for additive genetic variance, residual variance, and heritability (additive genetic variance-A, residual variance-B, heritability-C)

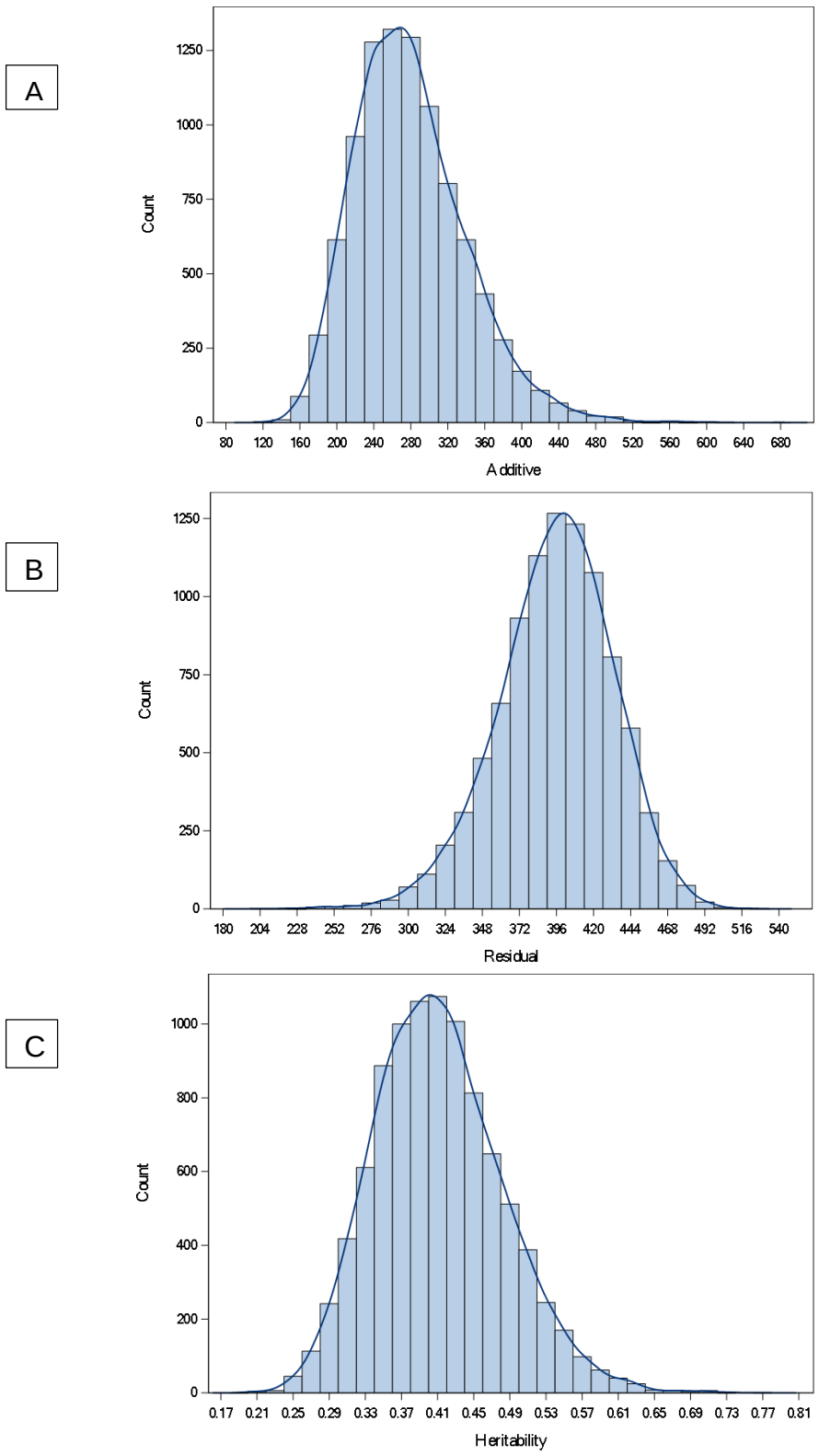



Figure 2: Posterior distribution of additive genetic variance, residual variance, and heritability from MCMC sampler (additive genetic variance-A, residual variance-B, heritability-C

### 4.6.4 Resampling methods

The estimates of variance components and corresponding heritability were estimated using four different resampling methods namely Non-Parametric Bootstrap, Parametric Bootstrap, Jackknife, and Asymptotic resampling method.

#### 4.6.4.1 Bootstrap estimation

##### 4.6.4.1.1 Non-parametric bootstrap (NPB) estimation

The distribution of the variance components and heritabilities obtained from non-parametric bootstrap is illustrated in Fig 3A-C. A total of 10,000 bootstrap samples were obtained, and for each sample, the additive genetic and residual variances were estimated by the REML method. Further, the expected values (mean of estimates) of the additive genetic variance (240), residual variance (230), and heritability (0.44 ± 0.05) over the bootstrap samples were obtained (Table 10). The 95 % confidence interval was constructed for variance components and heritability by calculating the $2.5^{th}$ and $97.5^{th}$ percentile of bootstrap distribution and is given in Table 10.

##### 4.6.4.1.2 Parametric bootstrap (PB) estimation

Under the parametric version of bootstrap, resampling was performed from a multivariate normal distribution with a mean zero and covariance matrix as specified for all random effects (additive and residual). The results obtained from the parametric bootstrap are presented in Table 10. The distribution of variance components and heritabilities are illustrated in Fig 4A-C. The estimates of variance components were obtained as the mean of the parameter estimates from bootstrap samples and the confidence interval as the values lying between $2.5^{th}$ and $97.5^{th}$ percentile. The estimated additive genetic variance from the parametric bootstrap was 255.6, and the residual variance was 325.2 (Table 10). The heritability estimated from the parametric bootstrap was 0.44±0.07 (Table 10).

#### 4.6.4.2 Asymptotic sampling

Another strategy used to approximate standard errors of heritability was through asymptotic sampling. The sampling was performed from a multivariate normal distribution parameterized with the mean and covariance matrix obtained from the REML analysis. The estimate of variance components was obtained as the mean of all the sampled values and is presented in Table 10. The distribution of the sampled values for additive genetic effects and residual effects are given in the Fig 5A and 5B. Estimates of heritability were obtained for each sampled value, and its distribution is presented in the Fig 5C. The heritability was estimated as the expected value (mean) across the asymptotic distribution and was 0.44 ± 0.07, which is the same as the REML estimate.

#### 4.6.4.3 Jackknife estimation

The total number of jackknife samples estimated for variance components and heritabilities was 1413. The distribution of variance components and heritabilities were highly asymmetrical (Fig 6A-C). The mean, mode, and median values of the distribution are given in Table 10. The mode (247.8) of additive genetic variance was similar to its mean (242.4) and median (243.5), and the mode (331.8), mean (337.4), and median (336.8) of residual variance were also similar. The heritability scores were similar between mean (0.42±0.01), median (0.42), and mode (0.43) obtained from the distribution.

### 4.6.5 Comparison of estimates from different methods

Variance components, heritabilities, and uncertainties were estimated from seven different methods as in ANOVA, REML, Bayesian method, non-parametric bootstrap, parametric bootstrap, jackknife and asymptotic sampling, the values of heritability ranged between 0.41 and 0.51 (Table 9-10). Estimates of heritability and its standard errors obtained from five methods viz., REML, parametric bootstrap, asymptotic sampling, the least square estimates of heritability obtained from the ANOVA model and the jackknife heritability were similar (0.44-0.45). MCMC sampling also gave a comparable heritability score of 0.41 (posterior mean) and 0.43 (posterior mode) with

an uncertainty of 0.07. The value of heritability estimated by the non-parametric bootstrap method was 0.51 ± 0.05 and was higher than the estimates from all other methods. Among the resampling methods, the jackknife method estimated the heritability with lowest standard error (0.42 ± 0.01). The 95% coverage probabilities of variance components and heritability from different methods are given in Table 11. The coverage probabilities obtained for different methods ranged from 0.94 to 0.98 (See Table 11). The coverage probabilities for the jackknife method were not estimated due to the small total sample size.

For subjective comparison, overlaid kernel density graphs for additive variance, residual variance, and heritabilities obtained from different sampling-based methods except for jackknife estimates are illustrated in Fig 7. These Figures show that the distribution of additive genetic variance, residual variance, and heritability is approximately normal. The location of the peak of distribution for the additive genetic variance doesn't vary among the methods. The smallest sampling variance for the additive genetic variance was obtained from non-parametric bootstrap (with a sharp peak and narrow confidence interval), giving higher confidence in the parameter estimated. Out of all the methods, the sampling variance for additive variance was high for the MCMC method with a wider confidence interval. The distribution for residual variance varied among methods (Fig 6), and a sharp peak with a narrow confidence interval was obtained for non-parametric bootstrap. Both the parametric bootstrap and asymptotic sampling gave rise to similar distribution for a residual variance. The residual variance obtained from the Bayesian posterior distribution was the highest in comparison to other methods. However, the sampling variance obtained from REML, parametric bootstrap, MCMC method, and asymptotic sampling for a residual variance were similar. The parametric bootstrap, MCMC method, and asymptotic sampling gave identical distribution for heritability estimates. A distinct sharp peak with a narrow confidence interval was observed in the non-parametric bootstrap method in contrast to other methods, which gave rise to a broader confidence interval.

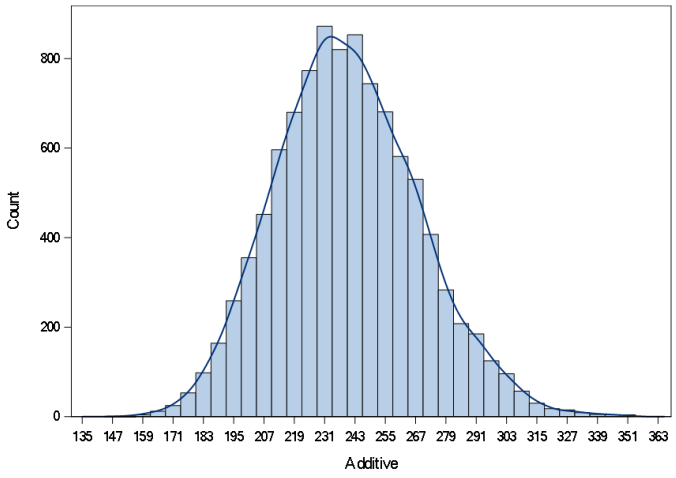
Table 10: Estimates of variance components, heritabilities with their sampling distributions and confidence intervals

| Method | Functions | Estimate | SE | L - 95% CI | U - 95% CI |
|---|---|---|---|---|---|
| **Non-Parametric Bootstrap** | Additive | 240.00 | 28.31 | 188.10 | 299.20 |
| | Residual | 230.00 | 24.00 | 182.60 | 277.10 |
| | Heritability | 0.51 | 0.05 | 0.41 | 0.61 |
| **Parametric Bootstrap** | Additive | 255.60 | 51.50 | 162.60 | 363.60 |
| | Residual | 325.20 | 31.90 | 259.70 | 384.50 |
| | Heritability | 0.44 | 0.07 | 0.30 | 0.58 |
| **Asymptotic Sampling** | Additive | 257.60 | 51.20 | 164.60 | 365.10 |
| | Residual | 325.70 | 31.60 | 266.30 | 389.00 |
| | Heritability | 0.44 | 0.07 | 0.30 | 0.58 |
| **Jackknife** | Additive | 242.48(Mean)<br>243.51(Median)<br>247.80(Mode) | 7.01 | 225.40 | 255.00 |
| | Residual | 337.40(Mean)<br>336.84(Median)<br>331.81(Mode) | 6.72 | 325.40 | 352.89 |
| | Heritability | 0.42(Mean)<br>0.42(Median)<br>0.43(Mode) | 0.01 | 0.39 | 0.44 |

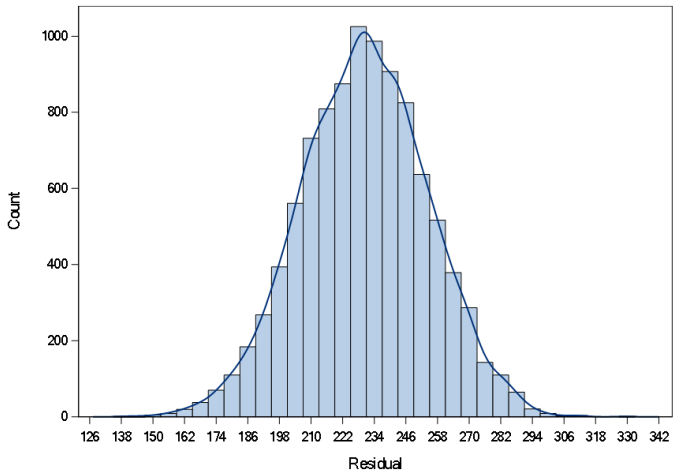
Table 11: Coverage probabilities for NPB: Non-parametric bootstrap, PB: Parametric bootstrap, BYS: Bayesian estimation, ASY: Asymptotic sampling

| | **NPB** | **PB** | **BYS** | **ASY** |
|---|---|---|---|---|
| **Additive** | 0.98 | 0.94 | 0.96 | 0.95 |
| **Residual** | 0.97 | 0.94 | 0.97 | 0.94 |
| **Heritability** | 0.97 | 0.94 | 0.96 | 0.95 |

A




B




C

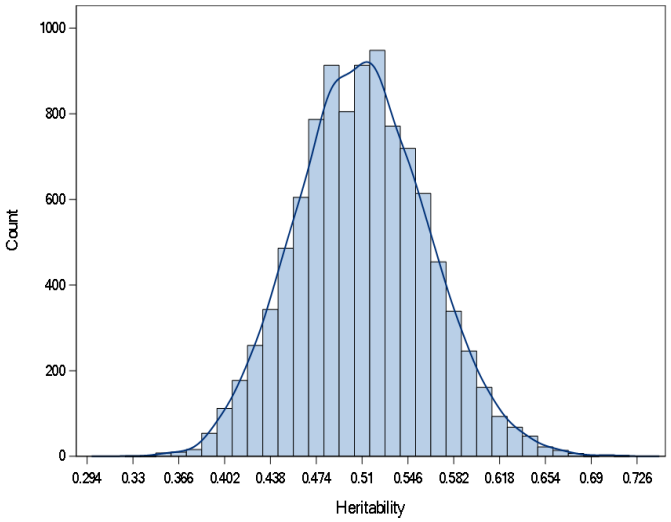
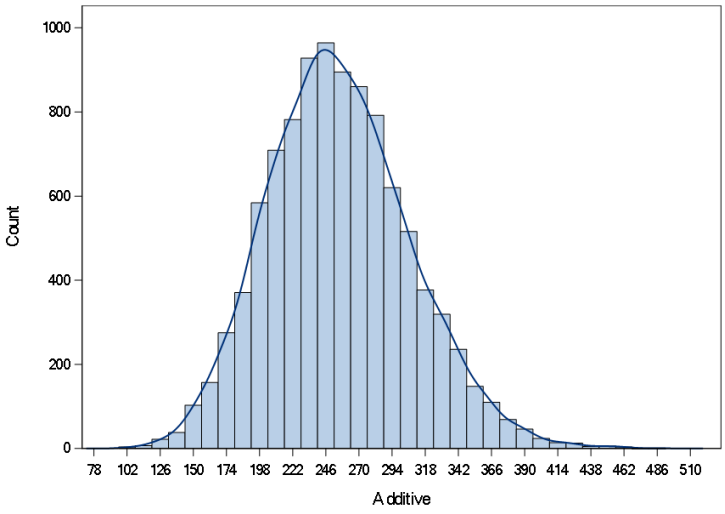
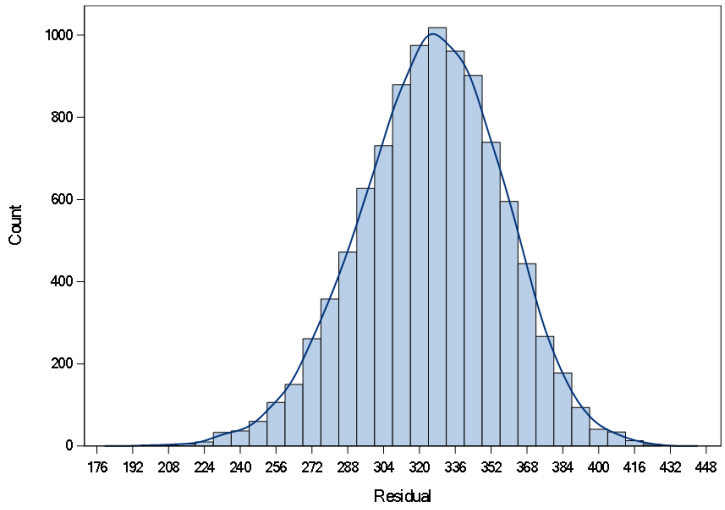



Figure 3: Distribution of additive genetic variance, residual variance, and heritability obtained from non-parametric bootstrapping (additive genetic variance-A, residual variance-B, heritability-C)

A




B




C

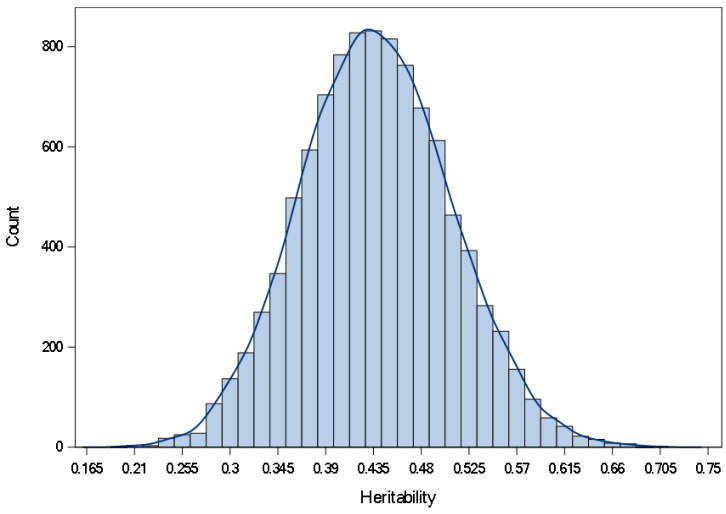
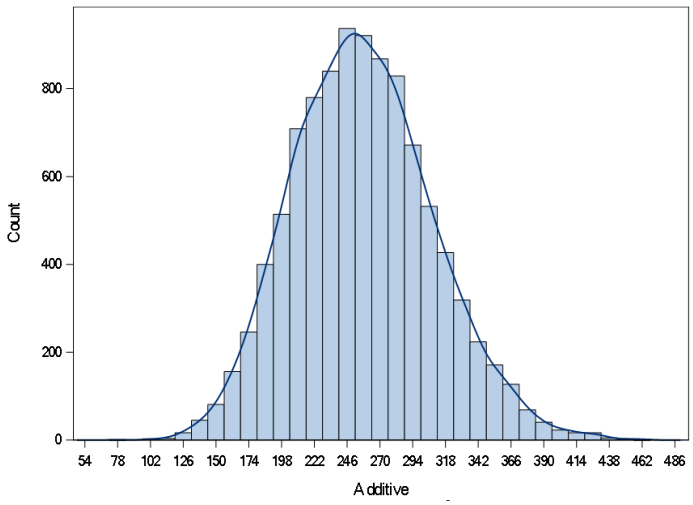
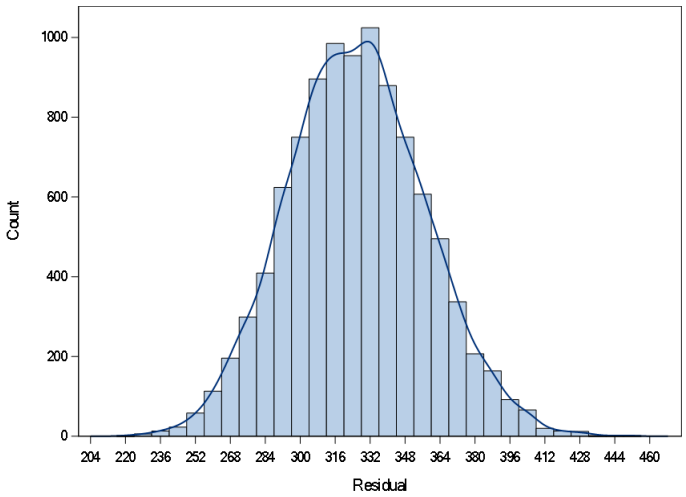



Figure 4: Distribution of additive genetic variance, residual variance, and heritability obtained from parametric bootstrapping (additive genetic variance-A, residual variance-B, heritability-C)

A




B




C

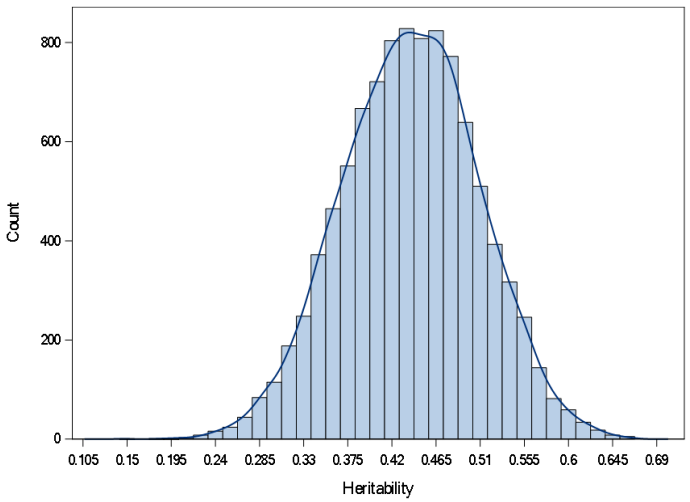
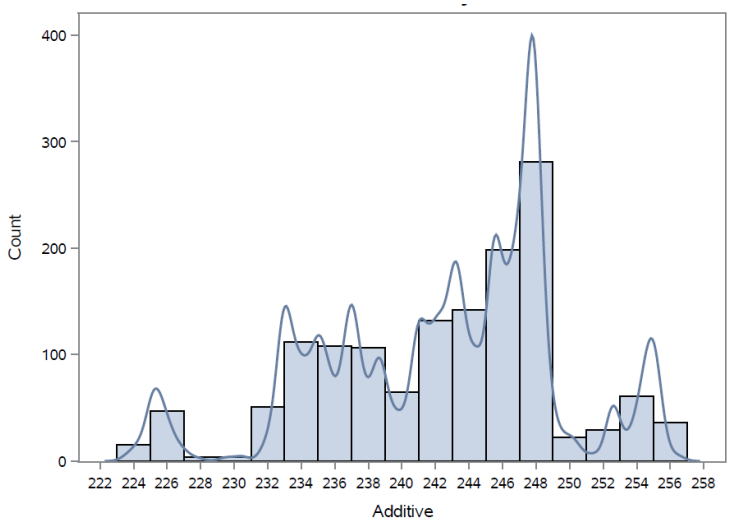
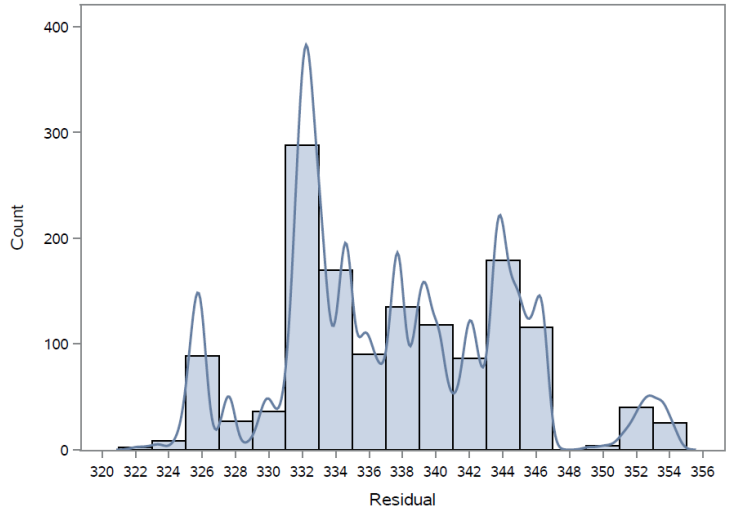



Figure 5: Distribution of additive genetic variance, residual variance and heritability obtained from asymptotic sampling (additive genetic variance-A, residual variance-B, heritability-C)

A




B




C

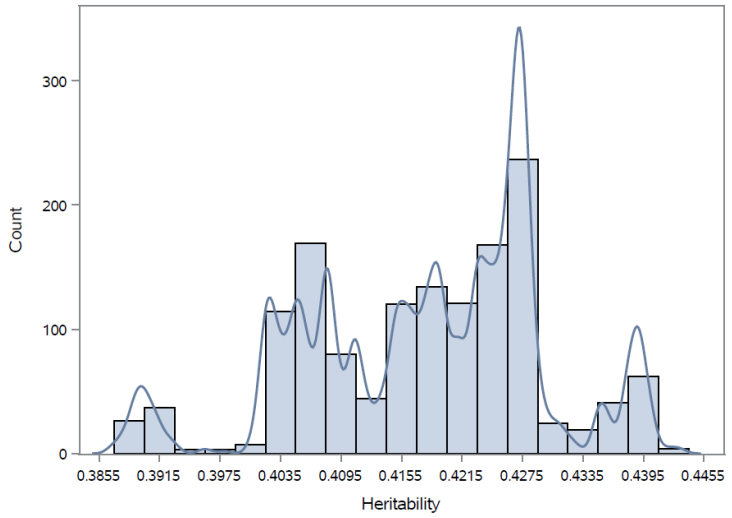
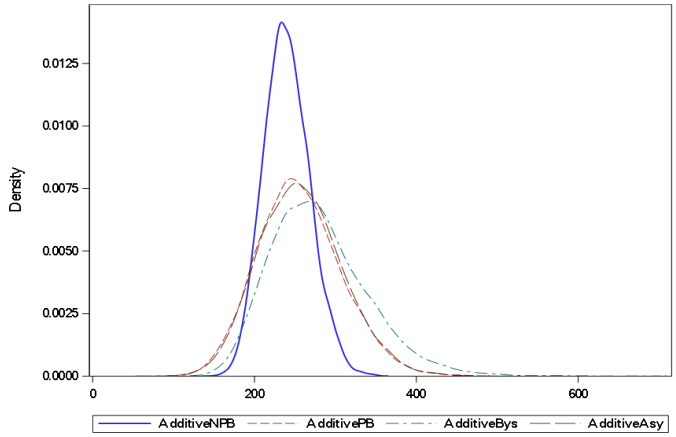
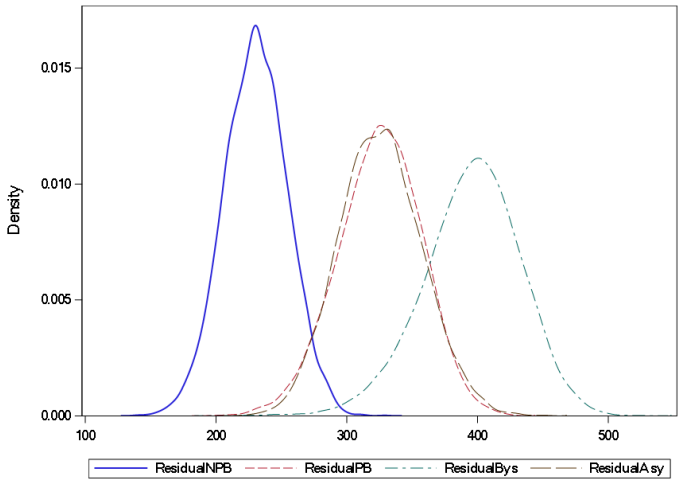



Figure 6: Distribution of additive genetic variance, residual variance and heritability obtained from jackknife sampling (additive genetic variance-A, residual variance-B, heritability-C

A




B




C

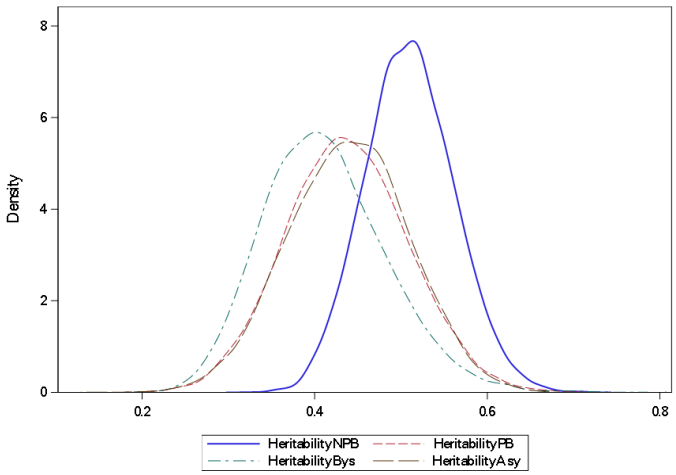



Figure 7: Overlaid kernel density graphs of additive genetic variance, residual variance, and heritability obtained from Non-Parametric Bootstrap (NPB), Parametric Bootstrap (PB), Bayesian (Bys), and Asymptotic sampling (Asy) methods ((additive genetic variance-A, residual variance-B, heritability-C)

### 4.7 Univariate and multivariate predictions of breeding values

The Breeding Values (BVs) for traits at harvest were predicted using both univariate Best Linear Unbiased Prediction (uBLUP) and multivariate Best Linear Unbiased Prediction (MBLUP) models. A full-sib model (FM) and an animal model (AM) were used to predict the family BVs and individual BVs. It was observed that there were differences in the predicted BVs obtained through univariate and multivariate models. The difference in predicted BVs obtained from univariate and multivariate models was captured as Spearman rank correlation values is presented in Table 12. The highest correlation value for both the models was observed for BW (0.99 for FM and 0.98 for AM), and the lowest rank correlation value was obtained for K (0.86 for FM and 0.80 for AM). The rank correlations between univariate and multivariate breeding values for BW, TL, BD, HW, and ADG were above 0.90 except for K (0.8 for AM and 0.86 for FM). The dumbbell plots (Fig 8-13) were used to visualize the differences in BVs predicted with univariate and multivariate models. The dumbbell plots for BD and K depicted a large difference between univariate and multivariate BVs compared to other traits.

#### 4.7.1 Accuracy of BVs from univariate and multivariate models

Prediction accuracies of family BVs and individual BVs from univariate and multivariate models for different traits were estimated and are depicted as line plots from Figures 14 to 18. There was an increase in the accuracy of BVs when predicted from multivariate models compared to the accuracy of BVs predicted from univariate models. The improvement in accuracy was less for family BVs in contrast to a more prominent increase of accuracy for individual BVs. Also, it was observed that there was a considerable increase in prediction accuracies for low heritable traits when a multivariate model was used compared to a modest increase of prediction accuracies for highly heritable traits. For instance, the increase in accuracy (average increase in accuracy by 13.1 % from parental model and 25.8 % from the animal model) of predicted BV was the highest for the trait K, which had the 0.07±0.02 heritability, and the change in accuracy was the lowest for the BW which had 0.44±0.07 heritability. The prediction accuracy of BVs for the parents was low in comparison to the offspring. Even though there was no

phenotypic records available for the parents, the animal model had predicted their BVs, of which the accuracies can be observed as irregularities at the beginning of line plots (Fig 14 to 18).

Table 12A: The Spearman rank correlations between the BVs obtained from univariate and multivariate predictions respectively for full-sib model (FM) and animal model (AM)

| Spearman Rank Correlation | | |
|---|---|---|
| **Traits** | **FM** | **AM** |
| **BW** | 0.99 | 0.98 |
| **TL** | 0.96 | 0.94 |
| **BD** | 0.93 | 0.91 |
| **HW** | 0.94 | 0.93 |
| **K** | 0.86 | 0.8 |
| **ADG** | 0.99 | 0.96 |

Table 12B: Absolute difference between the genetic and residual correlations of different traits obtained from the multivariate model

| **Traits** | **BW5** | **TL5** | **BD5** | **HW5** | **K5** | **ADG** |
|---|---|---|---|---|---|---|
| **BW5** | - | 0.25 | 0.52 | 0.19 | 0.49 | 0.04 |
| **TL5** | | - | 0.41 | 0.19 | 0.84 | 0.29 |
| **BD5** | | | - | 0.39 | 0.86 | 0.55 |
| **HW5** | | | | - | 0.64 | 0.24 |
| **K5** | | | | | - | 0.47 |
| **ADG** | | | | | | - |

Table 12C: Absolute values of genetic correlations (upper triangle) and absolute values of error correlations (lower triangle) between different traits obtained from the multivariate model

| Trait | BW5 | TL5 | BD5 | HW5 | K5 | ADG |
|---|---|---|---|---|---|---|
| **BW5** | - | 0.94 | 0.94 | 0.81 | 0.54 | 0.99 |
| **TL5** | 0.69 | - | 0.78 | 0.75 | 0.25 | 0.94 |
| **BD5** | 0.42 | 0.37 | - | 0.8 | 0.74 | 0.94 |
| **HW5** | 0.62 | 0.56 | 0.41 | - | 0.48 | 0.83 |
| **K5** | 0.05 | 0.59 | 0.12 | 0.16 | - | 0.52 |
| **ADG** | 0.95 | 0.65 | 0.39 | 0.59 | 0.05 | - |

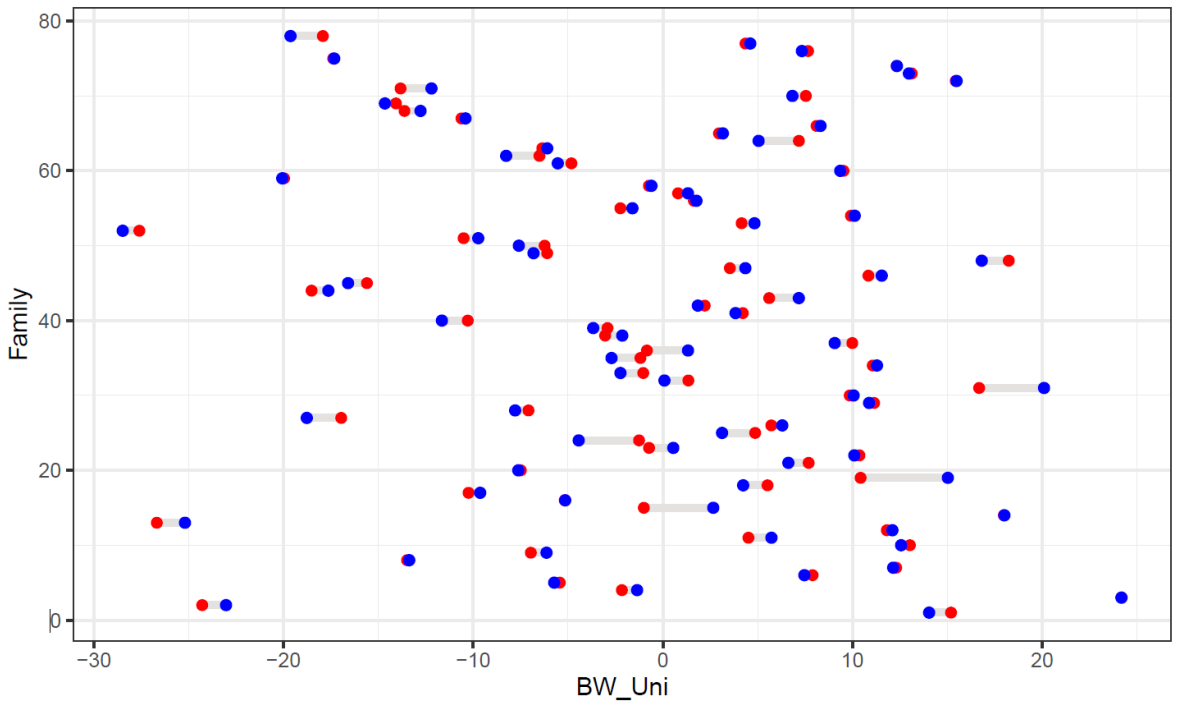



Figure 8: Dumbbell Plot depicting the difference in family BVs for BW at harvest between the Univariate (Red dots) and Multivariate (Blue dots) full-sib models

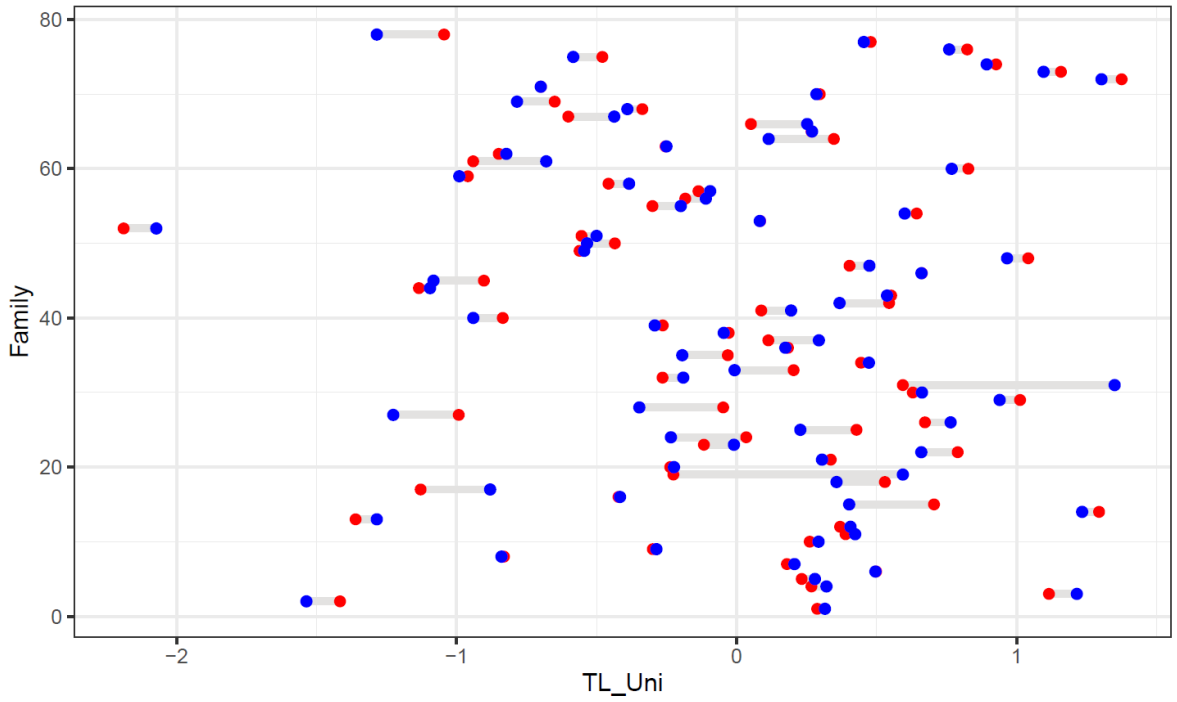



Figure 9: Dumbbell Plot depicting the difference in family BVs for TL at harvest between the Univariate (Red dots) and Multivariate (Blue dots) full-sib models

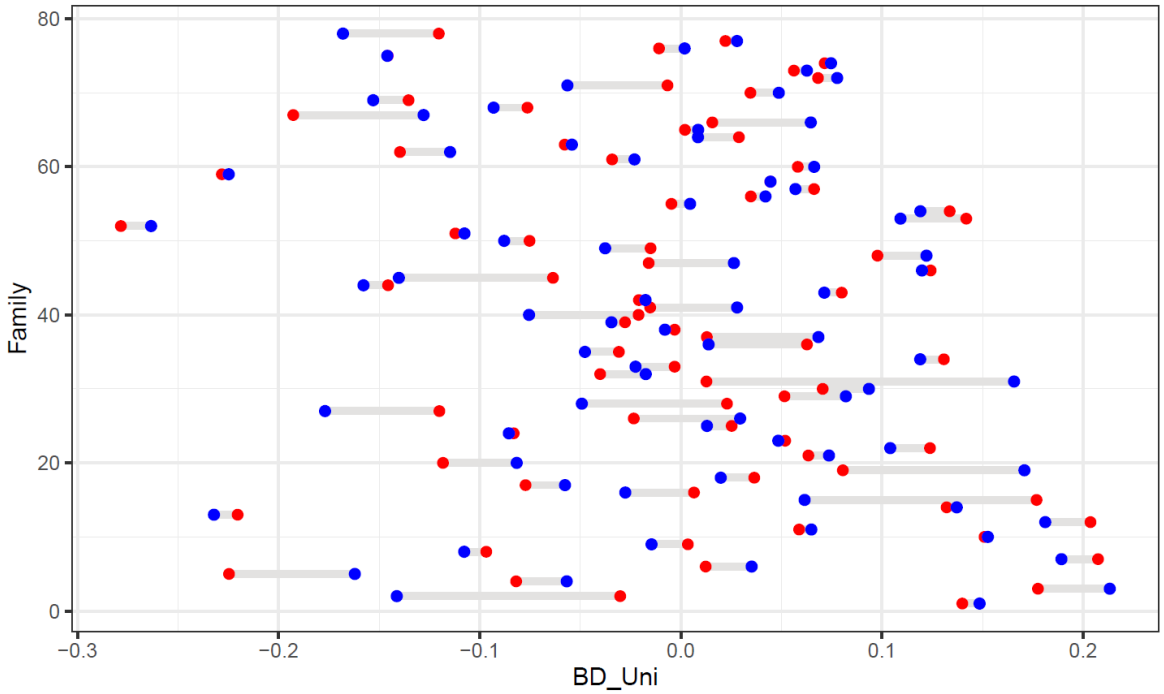



Figure 10: Dumbbell Plot depicting the difference in family BVs for BD at harvest between the Univariate (Red dots) and Multivariate (Blue dots) full-sib models

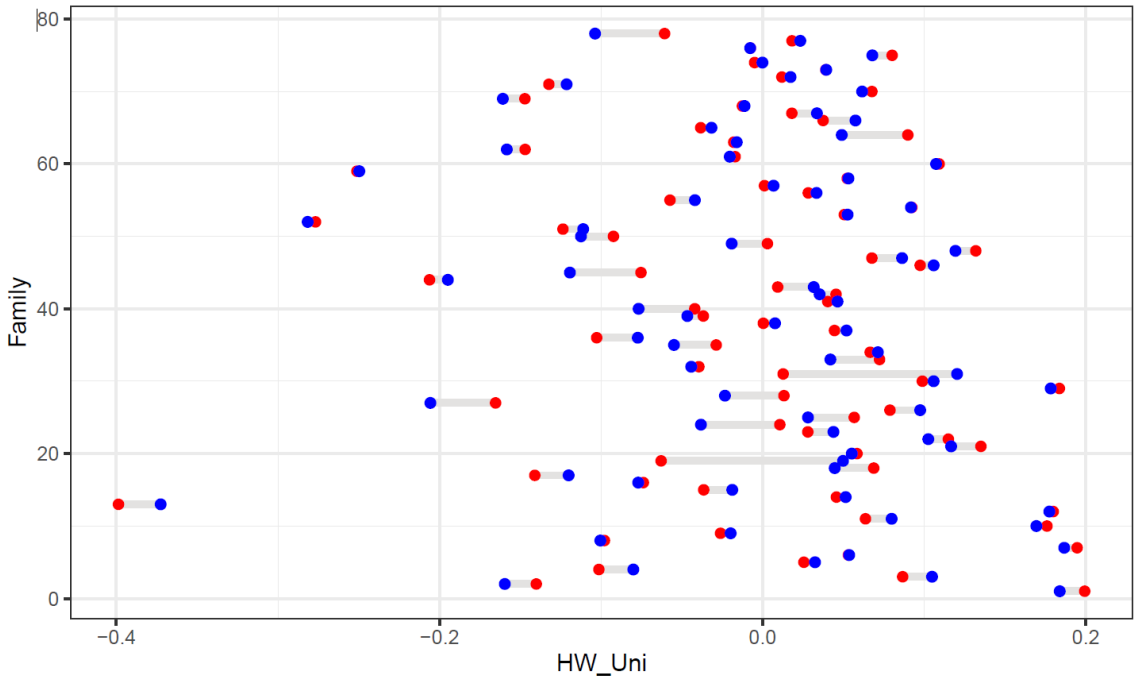



Figure 11: Dumbbell Plot depicting the difference in family BVs for HW at harvest between the Univariate (Red dots) and Multivariate (Blue dots) full-sib models

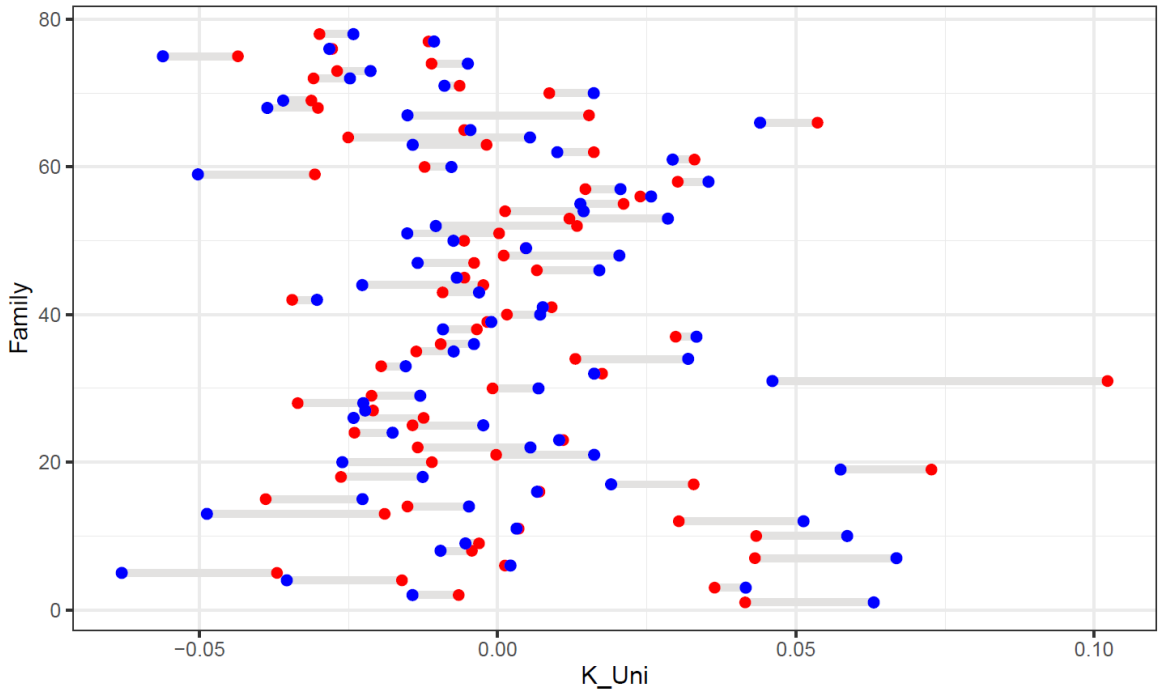



Figure 12: Dumbbell Plot depicting the difference in family BVs for K at harvest between the Univariate (Red dots) and Multivariate (Blue dots) full-sib models

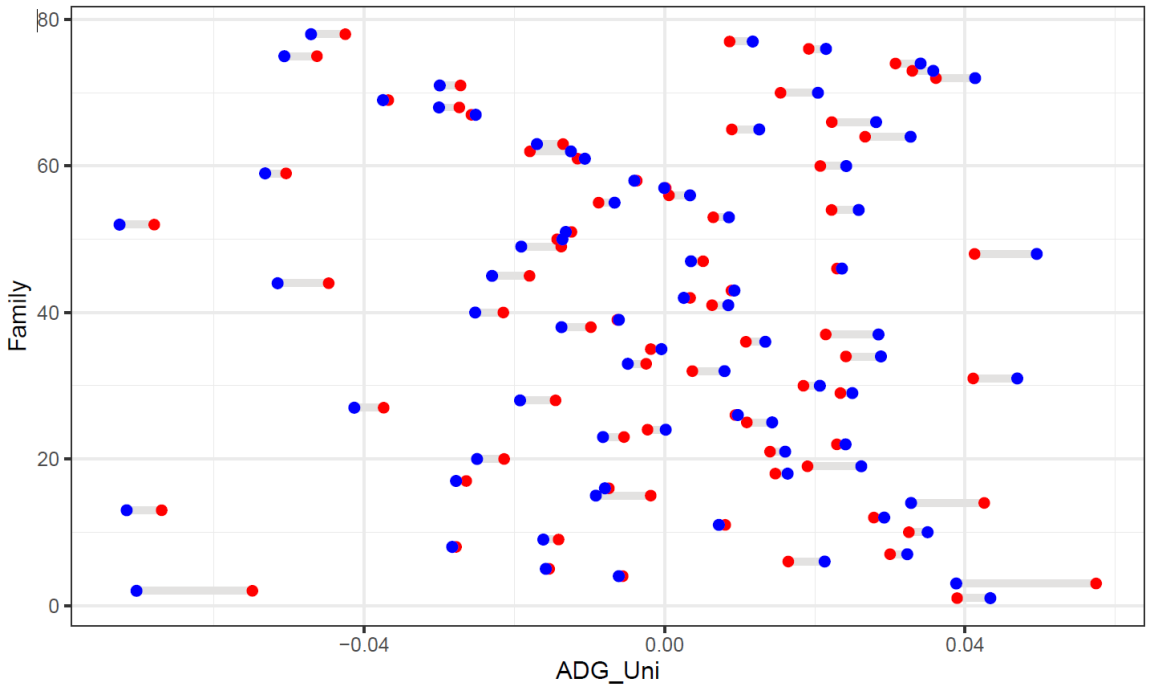



Figure 13: Dumbbell Plot depicting the difference in family BVs for ADG between the Univariate (Red dots) and Multivariate (Blue dots) full-sib models

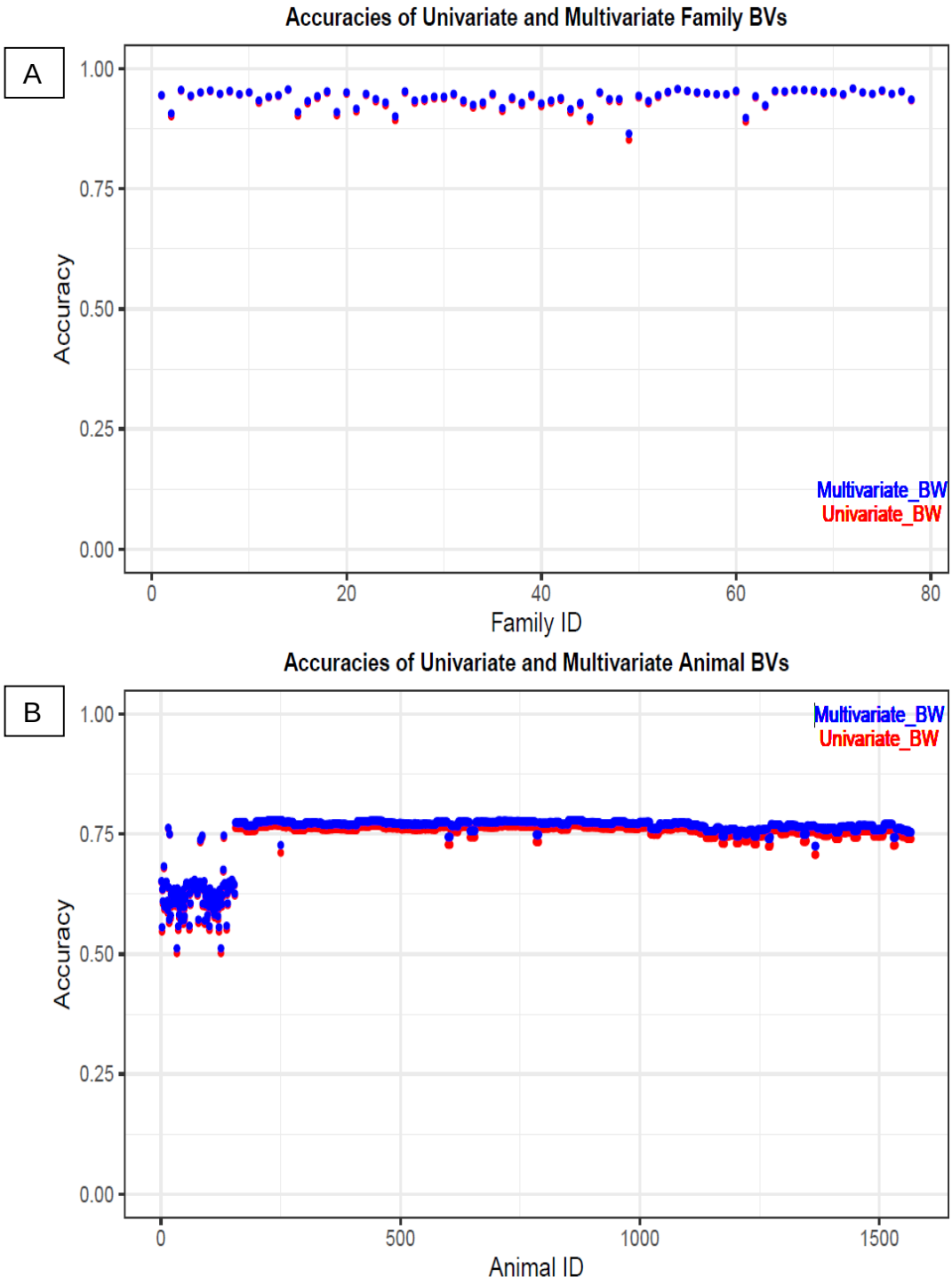



Figure 14: A - Plot depicting the difference in prediction accuracies of family breeding values for BW from univariate and multivariate models; B - Plot depicting the difference in prediction accuracy of individual breeding values for BW from univariate and multivariate models

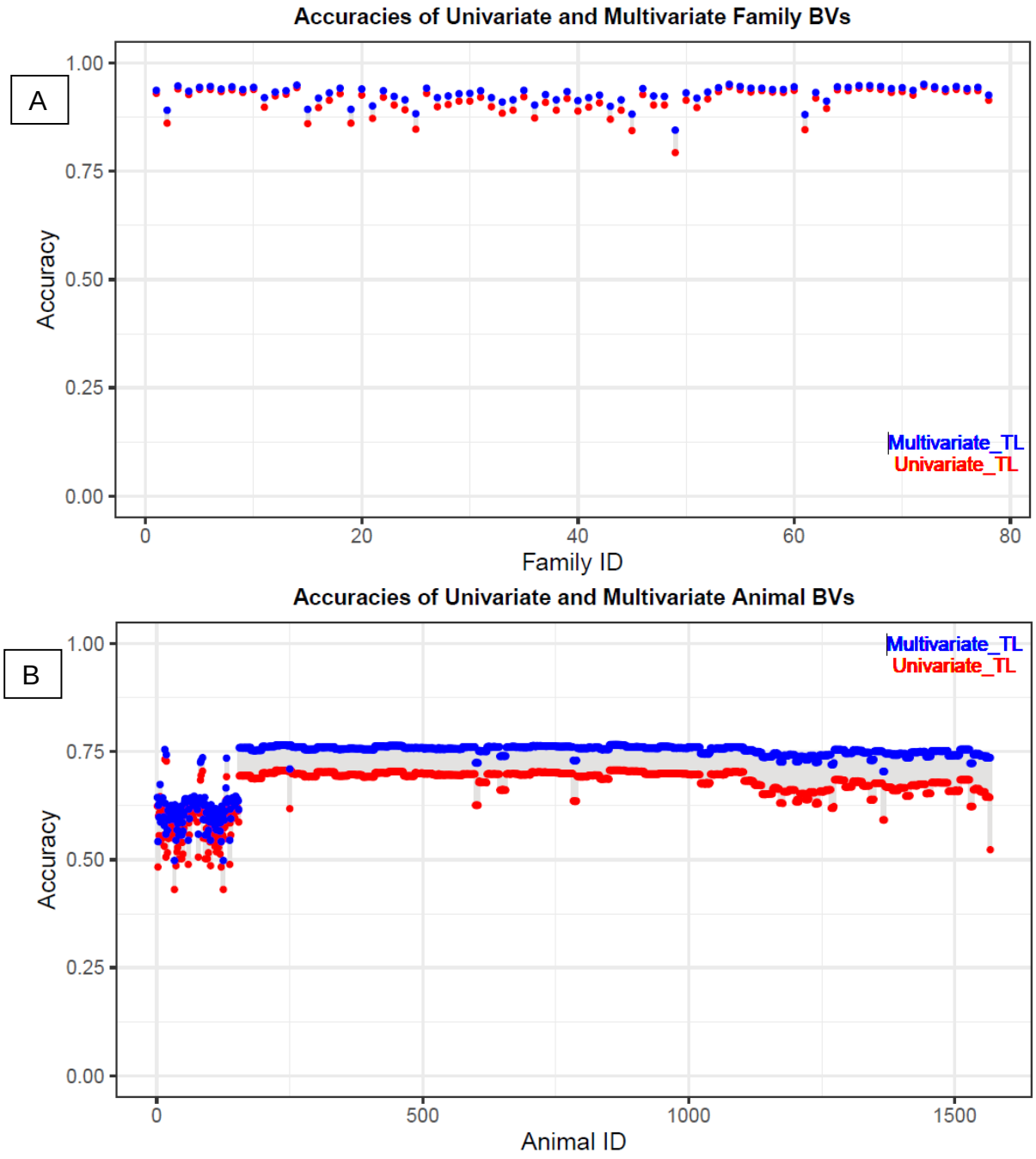



Figure 15: A - Plot depicting the difference in prediction accuracies of family breeding values for TL from univariate and multivariate models; B - Plot depicting the difference in prediction accuracy of individual breeding values for TL from univariate and multivariate models

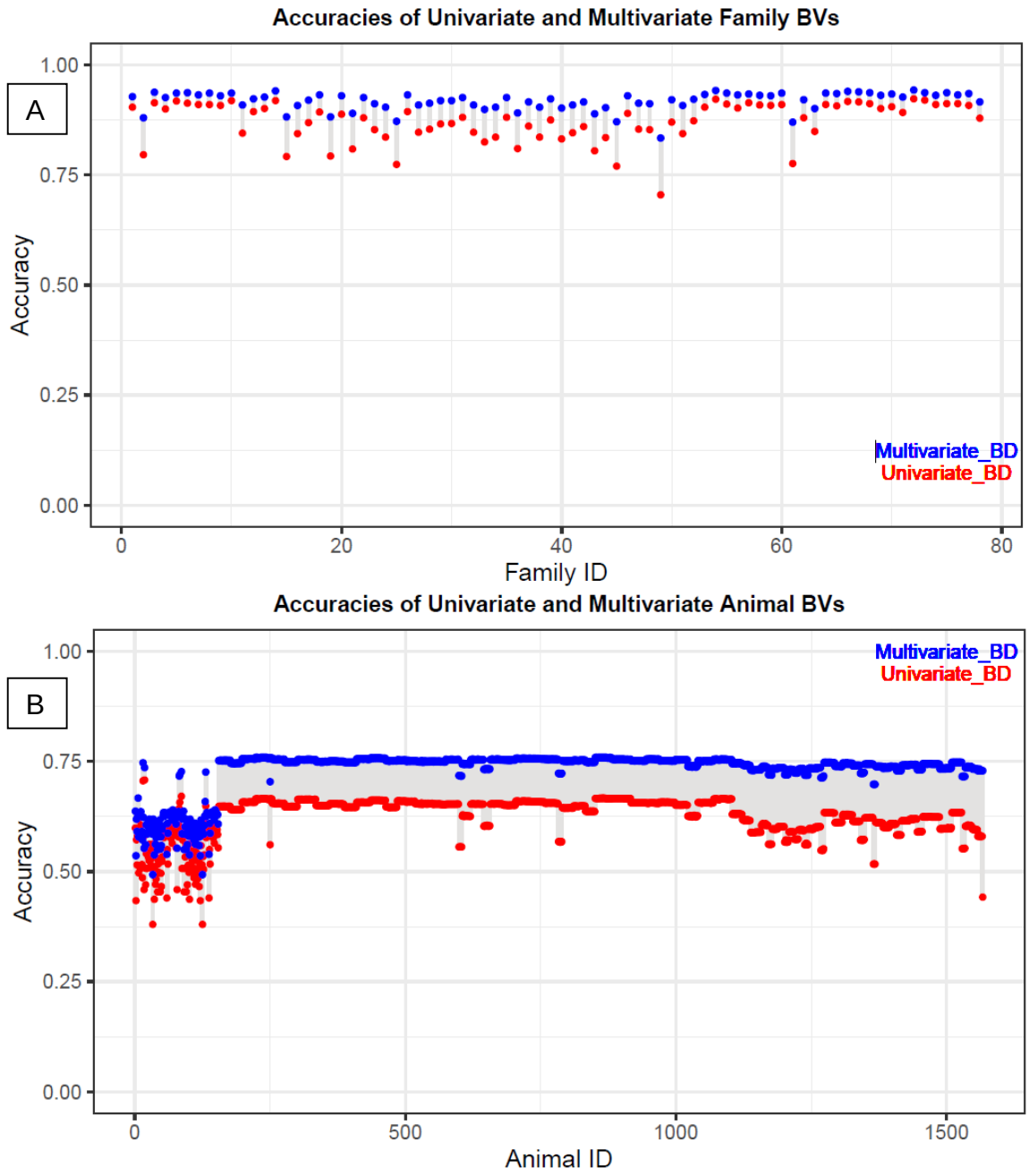



Figure 16: A - Plot depicting the difference in prediction accuracies of family breeding values for BD from univariate and multivariate models; B - Plot depicting the difference in prediction accuracy of individual breeding values for BD from univariate and multivariate models

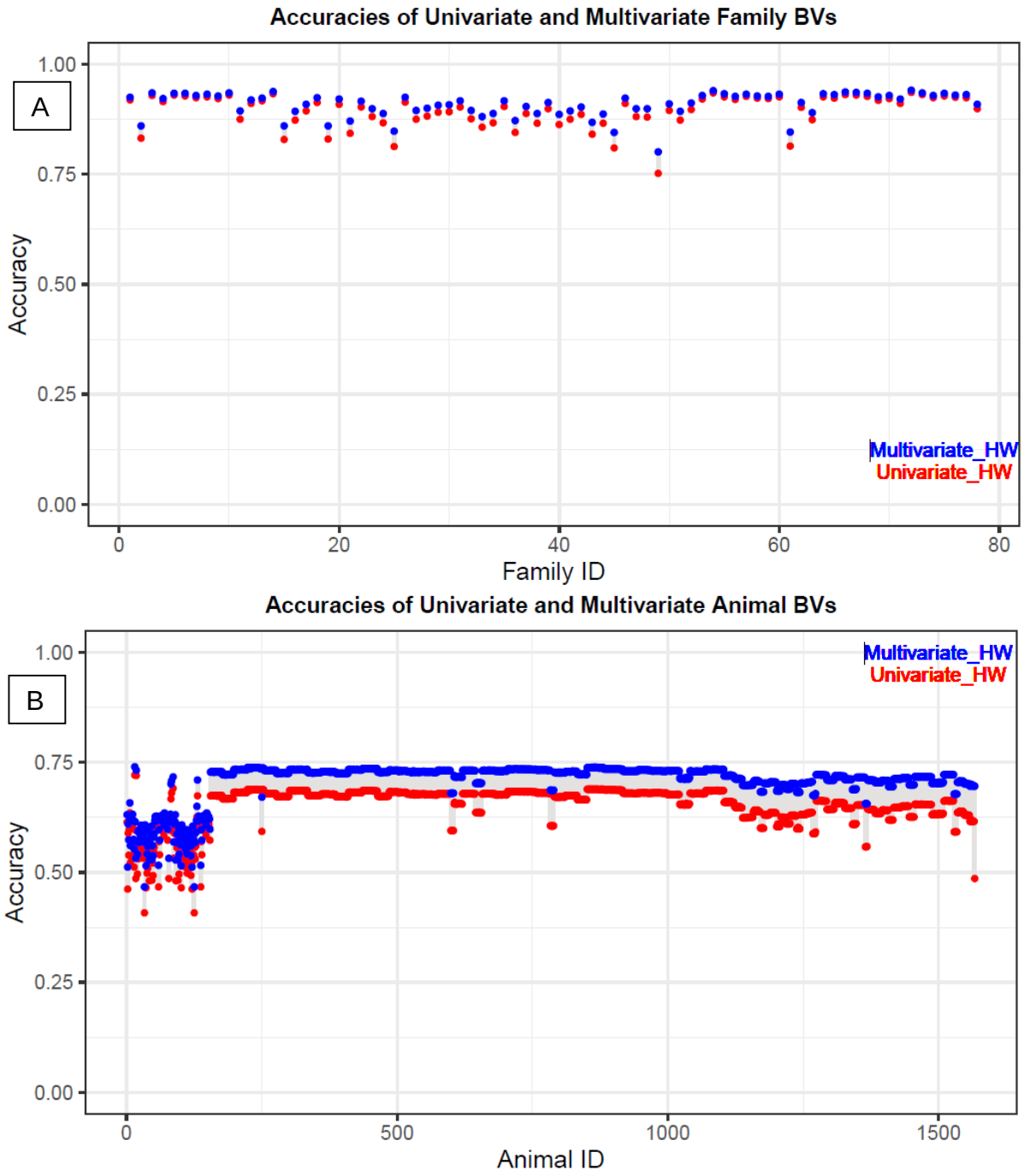



Figure 17: A - Plot depicting the difference in prediction accuracies of family breeding values for HW from univariate and multivariate models; B - Plot depicting the difference in prediction accuracy of individual breeding values for HW from univariate and multivariate models

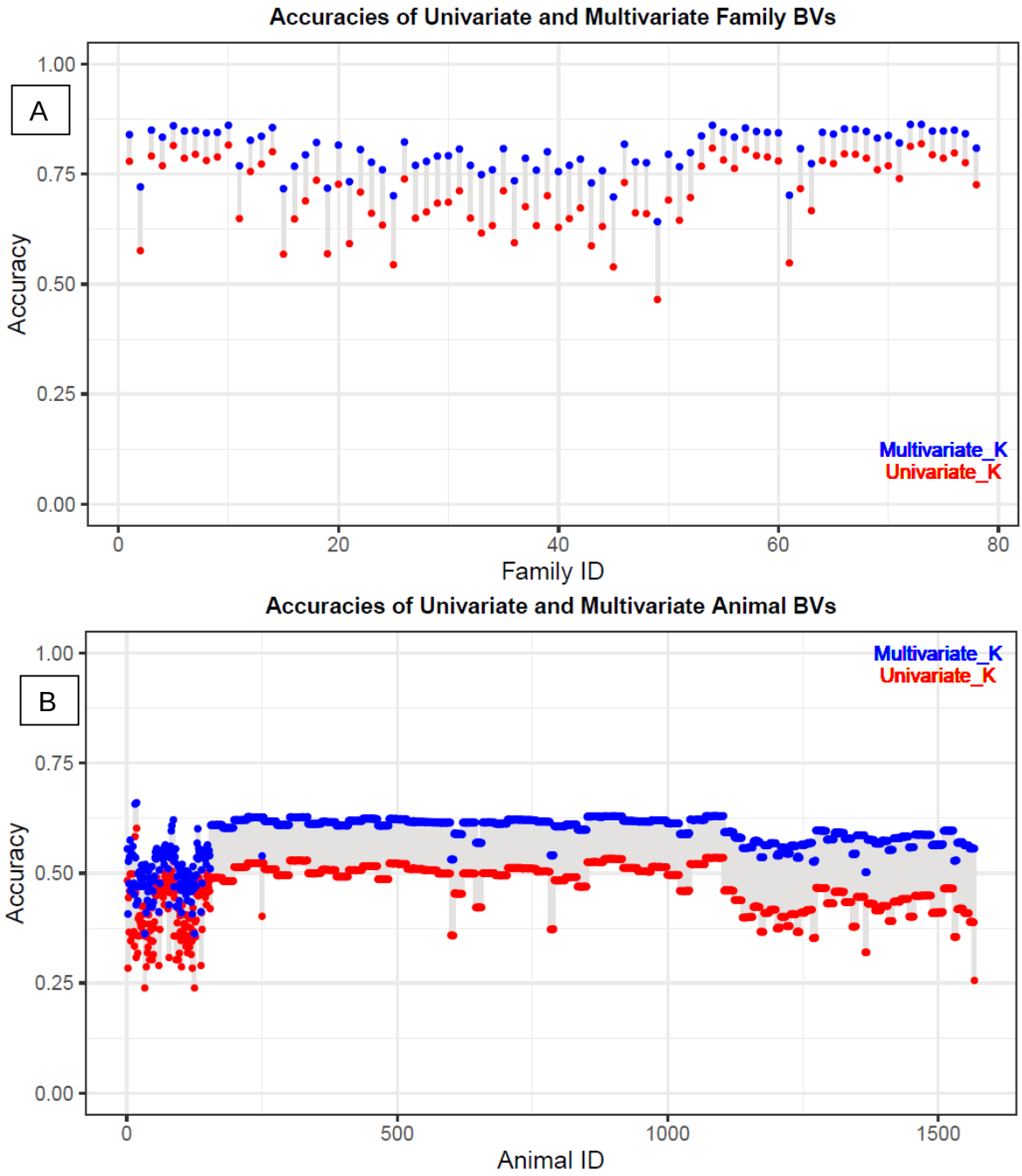



Figure 18: A - Plot depicting the difference in prediction accuracies of family breeding values for K from univariate and multivariate models; B - Plot depicting the difference in prediction accuracy of individual breeding values for K from univariate and multivariate models

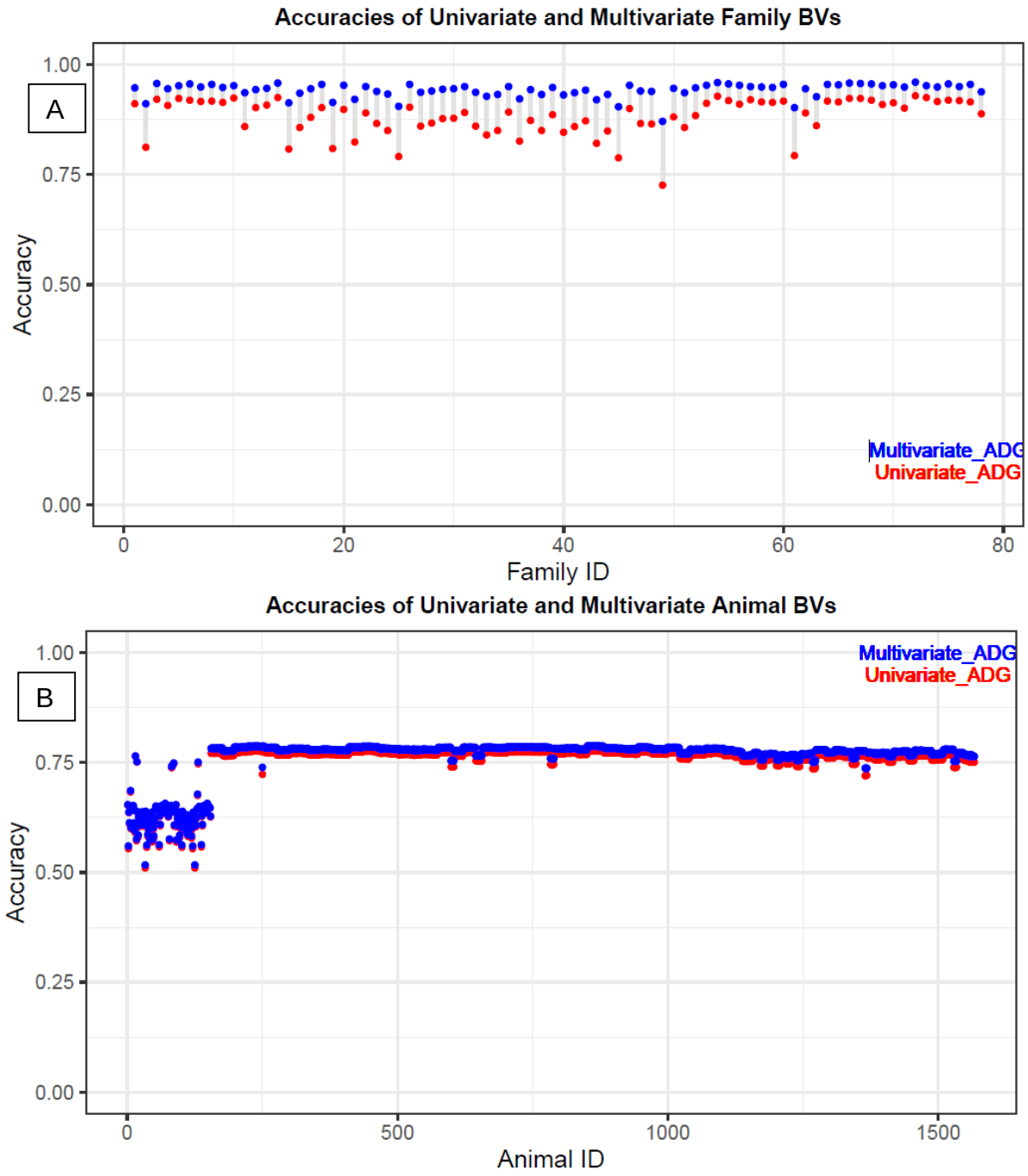



Figure 19: A - Line plot depicting the difference in prediction accuracies of family breeding values for ADG from univariate and multivariate models; B - Line plot depicting the difference in prediction accuracy of individual breeding values for ADG from univariate and multivariate models

## 4.8 Genetic evaluation of repeated measurements of body weight

### 4.8.1 Multi-trait model

#### 4.8.1.1 Heritability estimates

A multi-trait model was used to estimate the genetic parameters for the repeated measurements of body weight collected longitudinally at stocking (BW1), at three months (BW2), at six months (BW3), at nine months (BW4), and twelve months/harvest (BW5). The heritability, genetic and phenotypic correlation estimated from the multivariate model for (BW1-BW5) are presented in Table 13. The heritability had shown a decreasing trend over time with a very high value (0.84±0.09) at the time of stocking and high heritability (0.62±0.08) at harvest, but relatively lesser than at stocking. There were no conspicuous differences in heritability among body weight at six months (BW3: 0.64±0.08), nine months (BW4: 0.63±0.08) and twelve months/harvest (BW5: 0.62±0.08). The corresponding univariate heritability of BWs is also given in Table 13; the univariate heritability at each age was comparable to multivariate heritability. There was an increasing trend for permanent environmental effects from stocking to harvest and are presented in table 14. After the third sampling (BW3), the permanent environmental effect remains consistent till harvest. From the Table 14, an amount of variation corresponding to 19 percent at the time of harvest can be attributed to permanent environmental effect of the animal.

#### 4.8.1.2 Genetic and phenotypic correlations

The genetic and phenotypic correlations between the BWs were all positive and high. The genetic correlation between the body weight at stocking (BW1) and the rest of the traits (BW2-BW5) decreased as the sampling interval increased. The lowest genetic correlation (0.45±0.1) was observed between BW1 and BW5. The genetic correlation between the BWs measured from the adjacent intervals was high in comparison to the correlations measured from BWs far apart. The genetic correlation between the BWs recorded from three months till the harvest (BW2-BW5) were all high

(>0.90). A unit genetic correlation (1±0.01) was observed between BW4 and BW5 (Table 13). The phenotypic correlations between the traits BW1 to BW5 are also presented in Table 13. The phenotypic correlations revealed a similar trend as that of genetic correlations, with the lowest phenotypic correlation (0.47±0.04) observed between the BW1 and BW5. The phenotypic correlation was high between the BWs measured at the nearby intervals, and as the proximity of the sampling interval increased, the phenotypic correlation decreased (Table 13). The highest phenotypic correlation (0.95±0) was found between BW4 and BW5. The permanent environmental correlations were high between the BWs recorded at a later stage, with the highest correlation of 0.87±0.14 was observed between BW4 and BW5 (Table 14).

Table 13: Heritabilities (diagonal elements), genetic correlation (upper triangle), and phenotypic correlations between the repeated measures of BW sampled at three months interval obtained from the multi-trait model; the corresponding univariate heritabilities can be seen along the diagonal within the parentheses

| Traits | BW1 | BW2 | BW3 | BW4 | BW5 |
|---|---|---|---|---|---|
| **BW1** | **0.84±0.09 (0.91±0.09)** | 0.78±0.05 | 0.68±0.07 | 0.61±0.08 | 0.56±0.09 |
| **BW2** | 0.64±0.03 | **0.7±0.09 (0.70±0.09)** | 0.98±0.01 | 0.93±0.02 | 0.89±0.03 |
| **BW3** | 0.54±0.03 | 0.89±0.01 | **0.64±0.08 (0.62±0.09)** | 0.98±0.01 | 0.95±0.02 |
| **BW4** | 0.5±0.04 | 0.84±0.01 | 0.9±0.01 | **0.63±0.08 (0.63±0.09)** | 1±0 |
| **BW5** | 0.49±0.04 | 0.81±0.02 | 0.86±0.01 | 0.95±0 | **0.62±0.08 (0.64±0.09)** |

Table 14: Permanent environmental effects (diagonal elements), permanent environmental correlation (upper triangle) between the repeated measures of BW sampled at every 3 months interval obtained from the multi-trait model

| Traits | BW1 | BW2 | BW3 | BW4 | BW5 |
|---|---|---|---|---|---|
| **BW1** | **0.07±0.01** | 0.26±0.63 | 0.16±-1 | 0.25±0.45 | 0.34±-1 |
| **BW2** | | **0.12±0.08** | 0.81±0.23 | 0.74±0.32 | 0.72±0.24 |
| **BW3** | | | **0.18±0.01** | 0.74±0.22 | 0.68±0.34 |
| **BW4** | | | | **0.19±0.07** | 0.87±0.14 |
| **BW5** | | | | | **0.19±0.01** |

### 4.8.2 Random regression (RR) analysis

#### 4.8.2.1 Growth trajectories

Trellis plots depicting the growth trajectories of individual animals and the mean growth trajectory at different ages of batch-1 and batch-2 magur are presented in the Figure 20. Variations were observed at the level of intercept (a vertical shift of growth trajectories above and below the mean trajectory), slope (a gradient along the trajectories which were linear in nature), and the curvature of the trajectory (the gradient corresponding to the quadratic or higher-order polynomials). Also, the heteroscedastic nature of the growth trajectories with the increasing age can be readily observed from the plot. Similar plots for the magur reared in different ponds and from different families are depicted in Figures 21-22, respectively. Different levels of variation for growth trajectories can be readily observed from the plots for ponds and families (Fig 21-22).

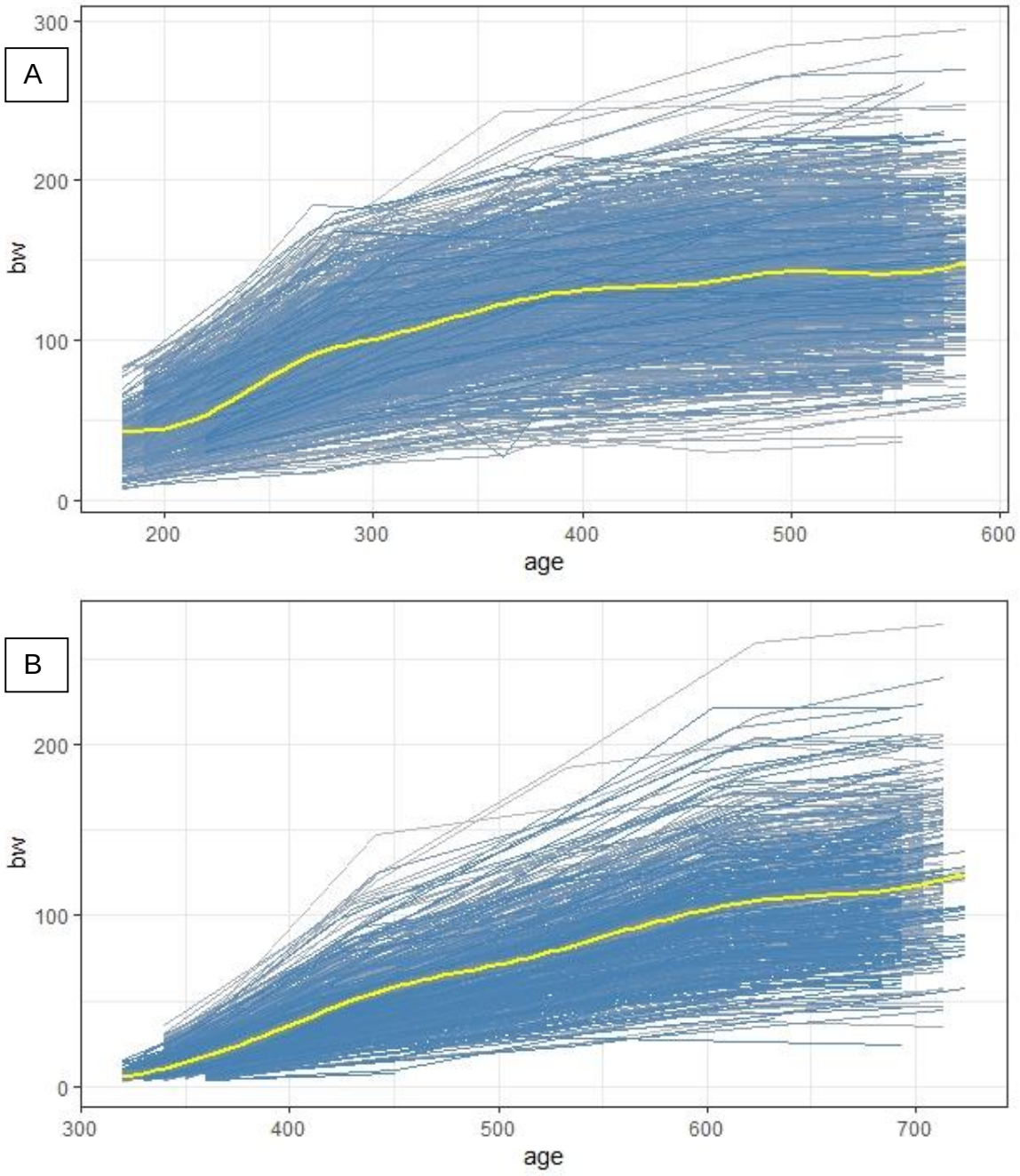



Figure 20: Trellis plot depicting the growth trajectories of magur along different ages- A-Batch-1; B-Batch-2

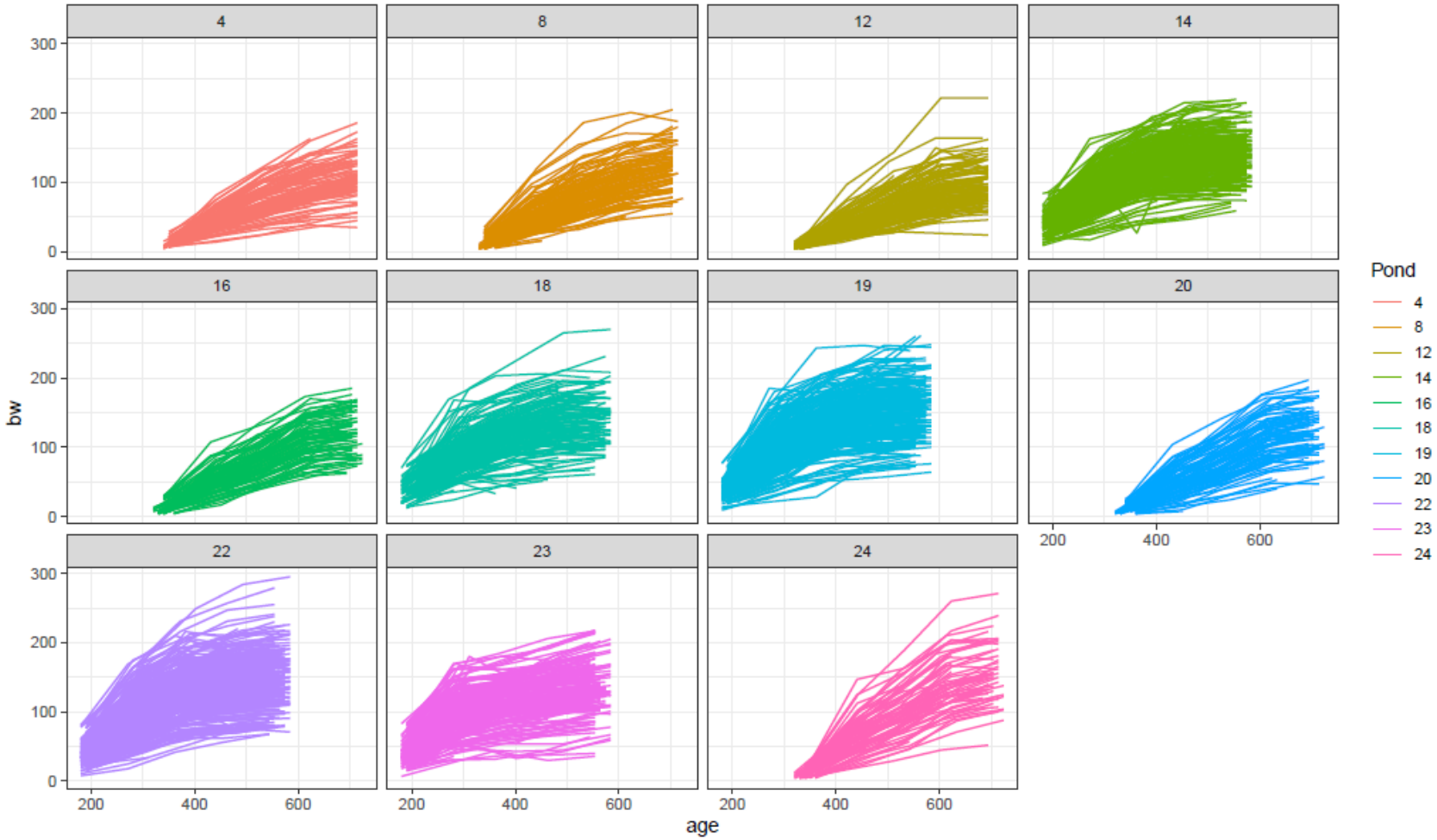



Figure 21: Trellis plot depicting growth trajectories of magur along different ages cultured in different ponds. The numbers on the upper strip are the identification numbers of the pond (4-24) and are labelled with colour codes. Pond Nos. 14, 18, 19, 22, 23 were from Batch-1 and Pond Nos. 4, 8, 12, 16, 20, 24 were from Batch-2

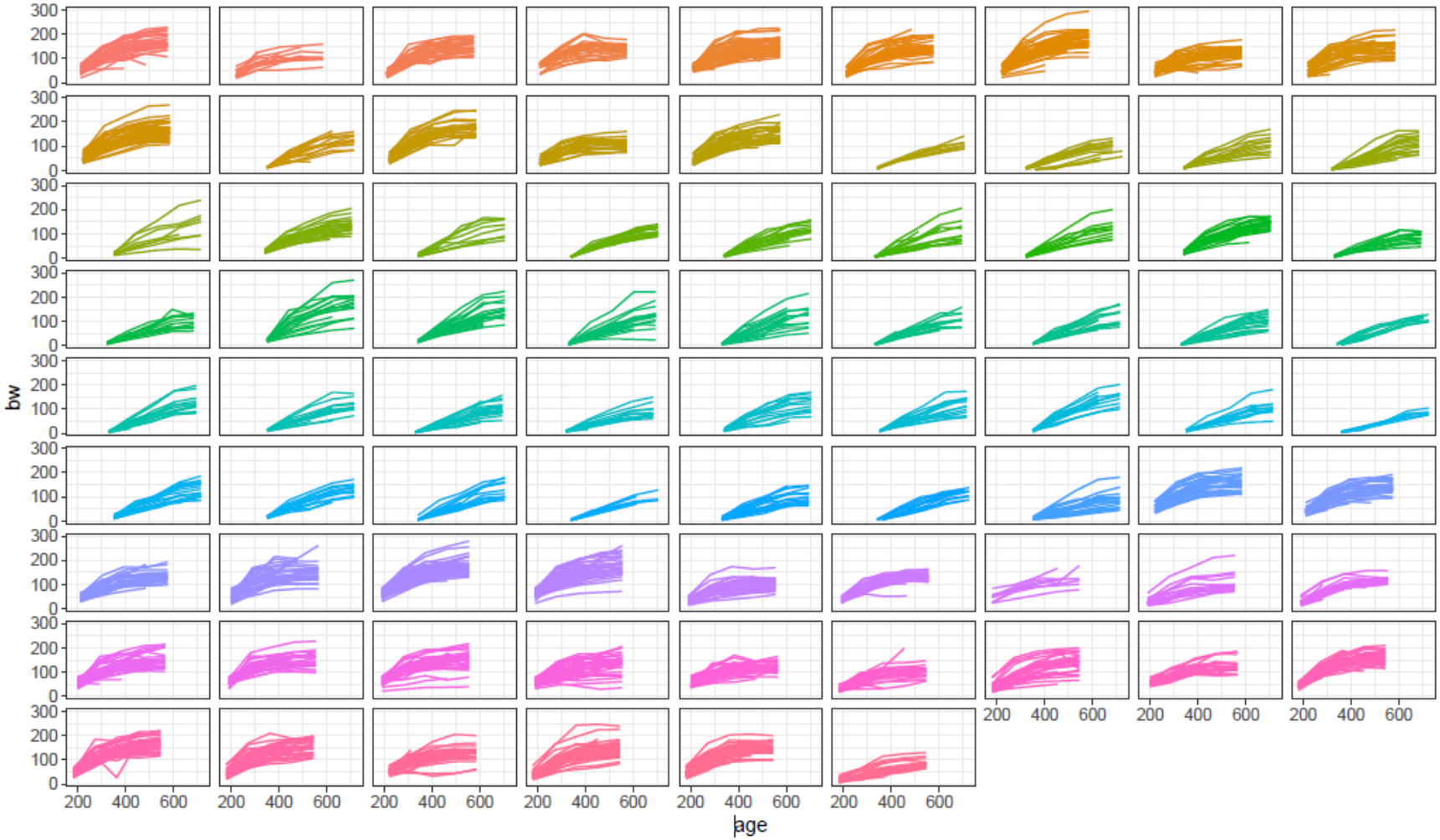



Figure 22: Trellis plot depicting the growth trajectories of magur from different families

#### 4.8.2.2 Preliminary analysis

The mean with standard error and the number of observations of BW measurements analyzed at 50 different ages across five sampling points from two batches are given in Table 15. At each sampling, the mean weights at different ages for batch-1 were higher than the mean weights of batch-2. The mean of BWs across ages ranged from 38.5±0.9 to 151±2.8 g for batch-1 between 180 and 584 days of age and 5.8±0.3 to 126.2±2.5 g for batch-2 between 320 and 724 days of age.

#### 4.8.2.3 Covariance functions

The reduced rank matrix of coefficients of the covariance functions corresponding to the intercept ($\boldsymbol{\beta_0}$), slope ($\boldsymbol{\beta_1}$), and quadratic ($\boldsymbol{\beta_2}$) terms of Legendre Polynomials (LP) for homogenous (HOM) and heterogeneous (HET) variance models are given in Table 16. In both HOM and HET models the additive genetic variance decreased as the order of polynomial increased. The drastic decrease in the variation along the axes (corresponding to intercept, slope, and quadratic terms) is captured as the percentage contribution of genetic eigenvalues. The genetic eigenvalue along the first axis (intercept) was many times higher in comparison to the eigenvalues of the other two axes (slope and quadratic). More than 85 % of variation of the trajectory was was accounted for by the intercept alone (Table 16). Hence, the majority of the genetic variation observed in the growth trajectory was resulted from the vertical shift of trajectory either above or below the mean trajectory. The percentage of genetic variation corresponding to the slope and curvature was 10.12 and 2.55 for HOM model and 8.9 and 2.77 for HET model. The genetic correlations between the quadratic terms of LP are given in the upper triangle of the covariance function matrix of Table 16. A positive genetic correlation (0.54±0.09) was observed between the intercept and slope, whereas the genetic correlation between the intercept and quadratic term, slope, and quadratic terms were negative.

The variance due to permanent environmental effects obtained from HOM and HET models can be seen in Table 16. The majority of the observed permanent environment variance (≈ 90%) was accumulated towards the vertical shifts in trajectories,

and a small amount of permanent environment variance (≈ 10 %) had a contribution towards the difference in slope of the trajectories (Table 16). However, there was no effect on the curvature of the slope due to permanent environment variance. The Bayesian Information Criteria (BIC) values and other parameters of the function of the BIC values are given in Table 16. The lowest BIC value was observed for the HET model, which suggests that, in this study, the HET model was better than the HOM model in explaining the variations in the growth trajectories.

Table 15: The mean and standard error (Mean±SE) of body weight (g) analyzed at 50 different ages across five different samplings (Stocking, 3, 6, 9, and 12 months) from two different batches. The rows represent measurements made on a similar age class in subsequent sampling

| Stocking | | | 3-Month | | | 6-Month | | | 9-Month | | | 12-Month | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Age** | **N** | **Mean±SE** | **Age** | **N** | **Mean±SE** | **Age** | **N** | **Mean±SE** | **Age** | **N** | **Mean±SE** | **Age** | **N** | **Mean±SE** |
| **180** | 260 | 38.5±0.9 | **271** | 238 | 92.2±2.0 | **362** | 234 | 122.5±2.3 | **453** | 235 | 135.3±2.5 | **544** | 214 | 141.8±2.6 |
| **190** | 277 | 48.5±1.1 | **281** | 258 | 97.8±2.1 | **372** | 241 | 128.2±2.2 | **463** | 229 | 137.9±2.6 | **554** | 219 | 142.9±2.8 |
| **200** | 116 | 55.7±1.4 | **291** | 92 | 101.5±3.0 | **382** | 107 | 124.3±2.8 | **473** | 101 | 133.6±3.2 | **564** | 89 | 137.0±3.5 |
| **210** | 281 | 45.3±0.9 | **301** | 263 | 98.2±1.6 | **392** | 253 | 130.1±1.8 | **483** | 250 | 138.3±2.0 | **574** | 239 | 143.8±2.0 |
| **220** | 253 | 50.1±1.0 | **311** | 235 | 104.1±1.9 | **402** | 216 | 132.5±2.4 | **493** | 209 | 147.1±2.6 | **584** | 191 | 151.0±2.8 |
| **320** | 72 | 7.6±0.30 | **411** | 47 | 38.9±1.6 | **502** | 48 | 70.0±2.9 | **593** | 54 | 97.6±4.0 | **684** | 58 | 106.8±3.9 |
| **330** | 157 | 6.2±0.20 | **421** | 119 | 41.9±1.4 | **512** | 117 | 67.6±2.1 | **603** | 124 | 99.4±2.9 | **694** | 116 | 108.4±3.3 |
| **340** | 192 | 12.8±0.6 | **431** | 144 | 56.4±1.6 | **522** | 140 | 84.7±2.2 | **613** | 153 | 112.2±2.5 | **704** | 145 | 126.2±2.5 |
| **350** | 163 | 13.4±0.4 | **441** | 113 | 60.7±2.2 | **532** | 124 | 86.1±3.1 | **623** | 135 | 114.0±3.4 | **714** | 129 | 124.9±3.6 |
| **360** | 28 | 5.8±0.30 | **451** | 18 | 24.0±2.3 | **542** | 17 | 50.4±3.4 | **633** | 22 | 81.1±4.4 | **724** | 19 | 98.4±4.7 |

Table 16: The estimated covariance matrix with standard deviations (in parenthesis) and corresponding Eigenvalues with the percentage contribution of the random regression coefficients. The coefficients are for the intercept ($\beta_0$), linear ($\beta_1$) and, quadratic ($\beta_2$), terms of Legendre polynomials for additive genetic and permanent environmental effects obtained from homogenous (HOM) and heterogeneous (HET) variance models; N –number of observations, P- number of parameter, Log- log likelihood and BIC-Bayesian Information criteria for HOM and HET Models.

| **Random Effects** | **Coefficient** | **HOM** | | | **HET** | | |
|---|---|---|---|---|---|---|---|
| | | $\boldsymbol{\beta_0}$ | $\boldsymbol{\beta_1}$ | $\boldsymbol{\beta_2}$ | $\boldsymbol{\beta_0}$ | $\boldsymbol{\beta_1}$ | $\boldsymbol{\beta_2}$ |
| **Additive Genetic** | $\boldsymbol{\beta_0}$ | **720.1 (131.9)** | 0.54 (0.09) | -0.72 (0.07) | **742.4 (135.7)** | 0.59 (0.08) | -0.70 (0.07) |
| | $\boldsymbol{\beta_1}$ | 165.1 (46.4) | **131.8 (24.2)** | -0.27 (0.13) | 184.8 (47.5) | **134.4 (24)** | -0.43 (0.12) |
| | $\boldsymbol{\beta_2}$ | -142.9 (31.3) | -22.5 (11.5) | **54.2 (10.4)** | -138.5 (30.5) | -36.4 (11.4) | **52.6 (9.8)** |
| | **Eigen Value** | 791.4 | 91.66 | 23.08 | 820.83 | 82.75 | 25.79 |
| | **%** | 87.34 | 10.12 | 2.55 | 88.32 | 8.9 | 2.77 |
| **Permanent Environment** | $\boldsymbol{\beta_0}$ | **451.8 (73.8)** | 0.69 (0.07) | -0.56 (0.11) | **480.8 (75.4)** | 0.80 (0.06) | -0.39 (0.12) |
| | $\boldsymbol{\beta_1}$ | 158.9 (27.9) | **115.9 (15.7)** | 0.21 (0.14) | 186.7 (28) | **114.5 (15.1)** | 0.25 (0.15) |
| | $\boldsymbol{\beta_2}$ | -62.1 (18.3) | 11.7 (7.6) | **27.3 (7.4)** | -40.5 (17.4) | 12.7 (7.1) | **22.6 (6.7)** |
| | **Eigen Value** | 520.84 | 74.16 | 0 | 561.15 | 56.76 | 0 |
| | **%** | 87.54 | 12.46 | 0 | 90.81 | 9.19 | 0 |
| **Model Criterion** | **N** | **P** | **Log** | **BIC** | **P** | **Log** | **BIC** |
| | 7754 | 13 | -25270 | 50464.49 | 62 | -24804.5 | 49659.61 |

#### 4.8.2.4 Variance estimates

The estimates of variance components at different ages constructed from the estimated covariance functions from different RRM are illustrated in Figures 23-27. The additive genetic variance tends to sharply increase with age till 500 days and subsequently drops (Fig 23). Both the models have shown a similar pattern of increase in additive genetic variance over time which was non-linear in nature. The variances due to permanent environmental effects tend to increase with time in a linear fashion (Fig 24). Both the RRM gave similar results for the permanent environmental variance. The phenotypic variance tend to increase steadily in HOM model and a steady increase with irregularities in the HET and HET-F models (Fig 25). There was no pattern for the residual variance estimated from the HET models (Fig 26 – 27).

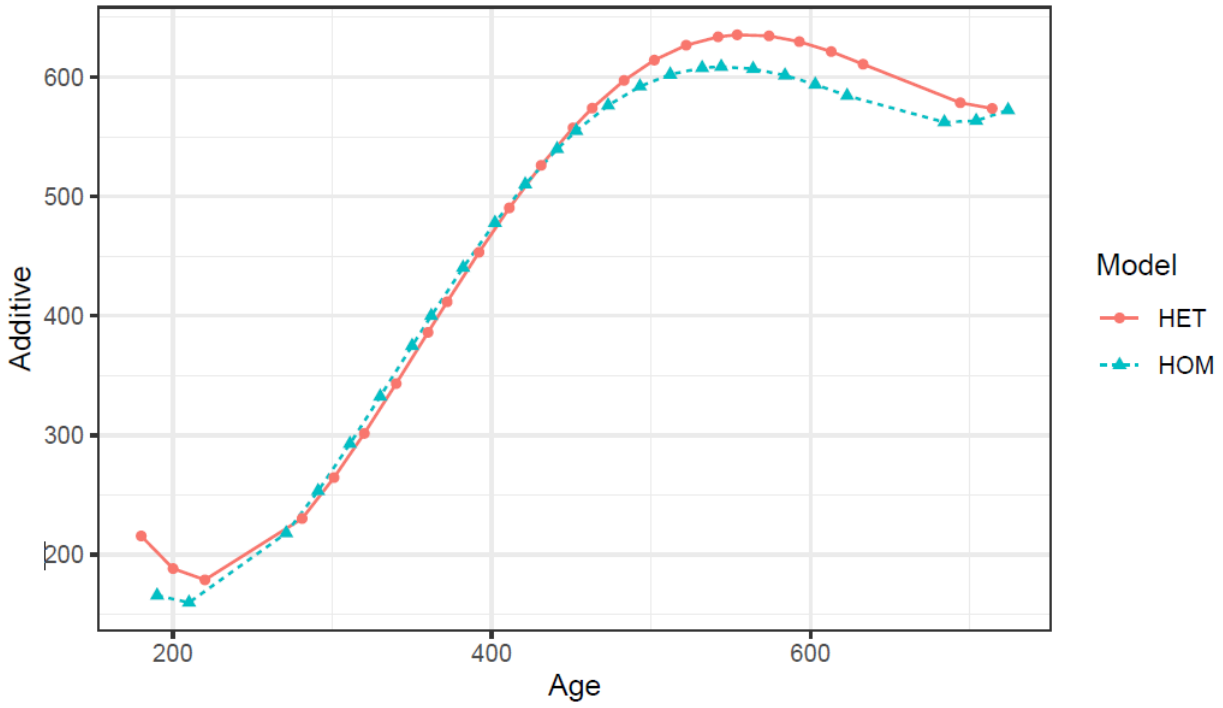



Figure 23: Plot representing the change in additive genetic variance at different ages estimated from homogeneous (HOM) and heterogeneous (HET) models

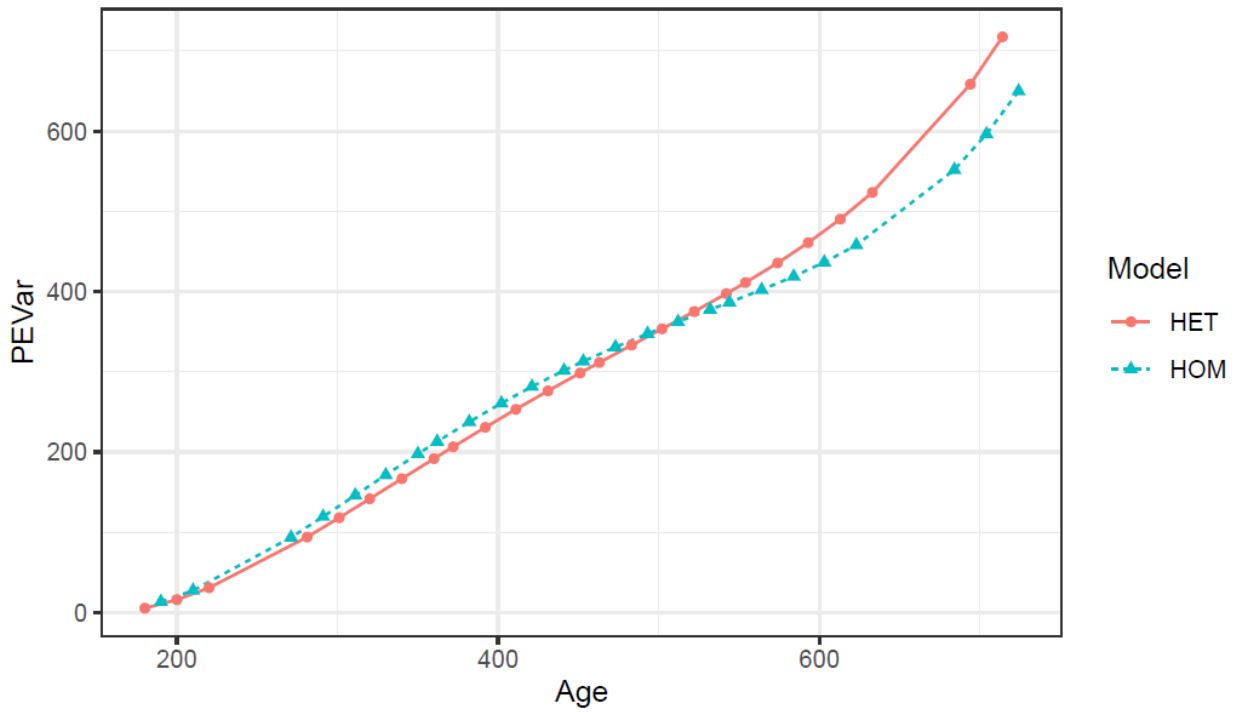



Figure 24: Plot representing the change in permanent environmental variance (PEVar) at different ages estimated from homogeneous (HOM) and heterogeneous (HET) models

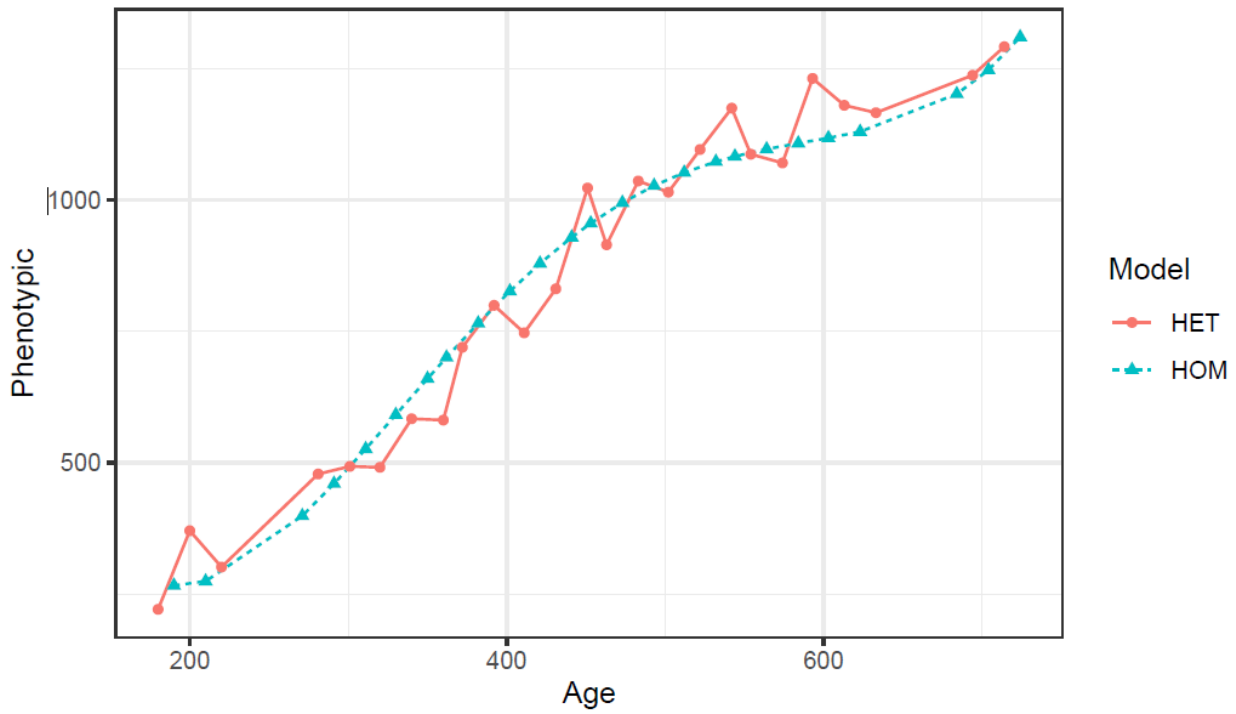



Figure 25: Plot representing the change in phenotypic variance at different ages estimated from homogeneous (HOM) and heterogeneous (HET) models

.

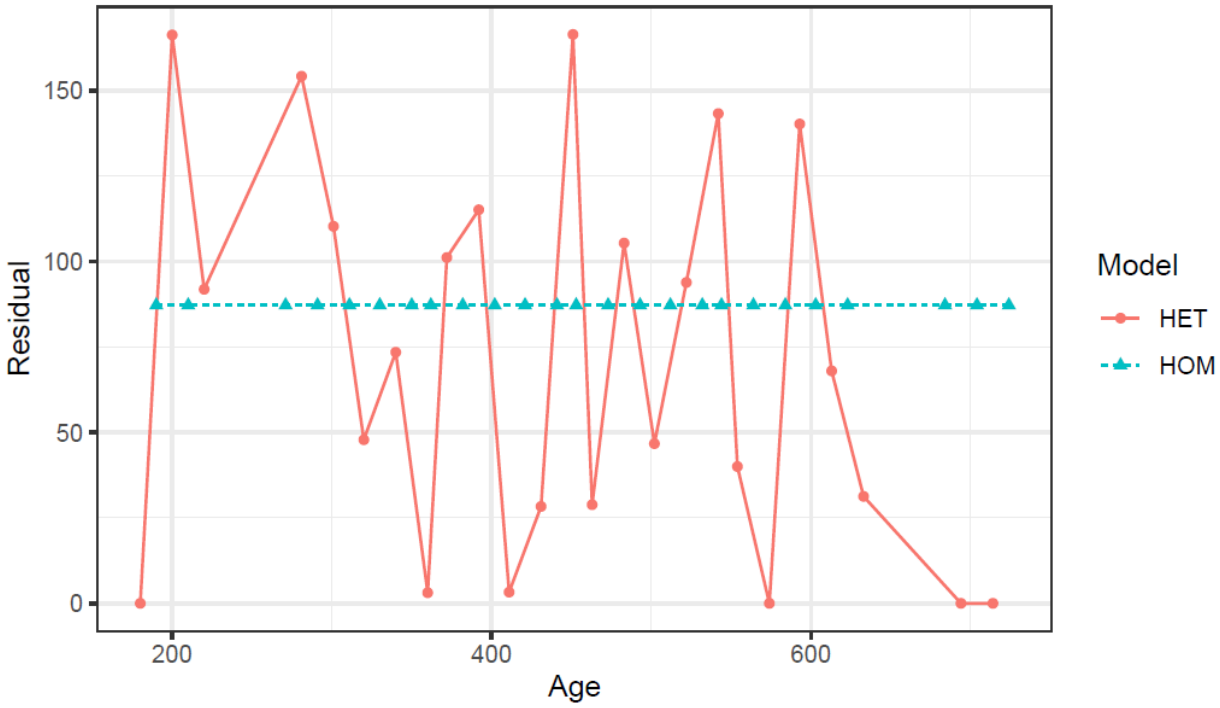



Figure 26: Plot representing the change residual variance at different ages estimated from homogeneous (HOM) and heterogeneous (HET) models

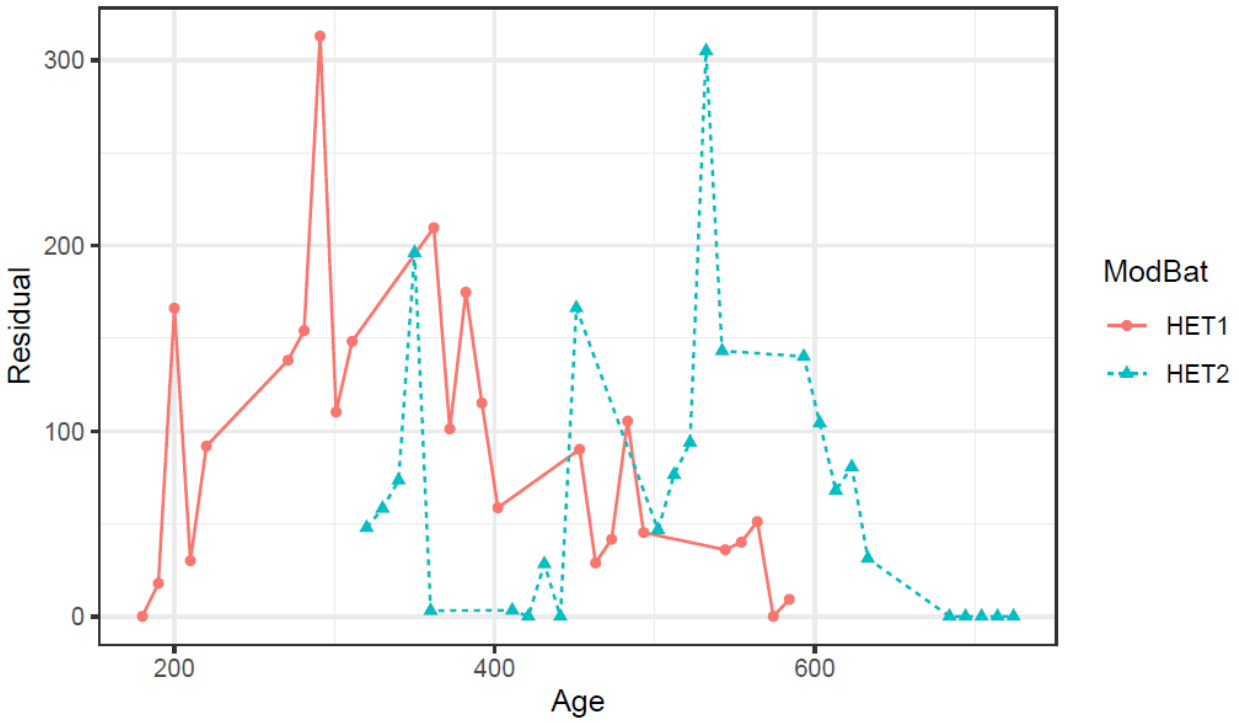



Figure 27: Plot depicting the batch wise change in residual variances at different ages estimated from heterogeneous (HET) models

#### 4.8.2.5 Heritability estimates

The heritability estimated at different ages from the estimated covariance function for HOM and HET RRM is depicted in Figure 28 (Table 17 A-C). The heritability ranged from 0.42 to 0.99 for HET model, and 0.43 to 0.65 for the HOM model. For both the models, the heritability first decreased with an increase in age. Over the ages, the heritabilities estimated from HET models showed a decreasing trend but with irregular patterns as opposed to a smooth decreasing age trend in HOM model heritability values. The heritability scores from HOM and HET models converged towards the final ages. Figure 29 and 30 represent the HET and HOM heritabilities by batch, which reveals different patterns for the heritabilities estimated by HET models whereas HOM model heritabilities followed a similar pattern acorss the batches.

The permanent environmental (PE) effects estimated from the RR models are depicted in Figure 31 and also listed in Tables 17 A-C. An increasing trend with age was observed for PE for all the models, the PE effect was almost nil at the beginning and then gradually increased to reach up to ≈ 0.50 at the time of harvest.

Table 17A: Heritability ($h^2$), permanent ($p^2$) and common ($c^2$) environment effects estimated at different ages from homogenous and heterogeneous variance model (HOM/HET)

| Age | HOM | | HET | |
|---|---|---|---|---|
| | $h^2$ | $p^2$ | $h^2$ | $p^2$ |
| **180** | 0.65±0.07 | 0.03±0.06 | 0.97±0.09 | 0.03±0.1 |
| **190** | 0.62±0.07 | 0.05±0.06 | 0.88±0.09 | 0.05±0.09 |
| **200** | 0.6±0.07 | 0.08±0.06 | 0.51±0.08 | 0.04±0.05 |
| **210** | 0.58±0.07 | 0.1±0.06 | 0.77±0.08 | 0.1±0.08 |
| **220** | 0.57±0.07 | 0.13±0.05 | 0.59±0.07 | 0.1±0.06 |
| **271** | 0.55±0.07 | 0.24±0.06 | 0.49±0.07 | 0.19±0.05 |
| **281** | 0.55±0.07 | 0.25±0.06 | 0.48±0.07 | 0.2±0.05 |
| **291** | 0.55±0.07 | 0.26±0.06 | 0.37±0.06 | 0.16±0.04 |
| **301** | 0.55±0.07 | 0.27±0.06 | 0.54±0.07 | 0.24±0.06 |
| **311** | 0.56±0.07 | 0.28±0.07 | 0.5±0.07 | 0.23±0.06 |
| **320** | 0.56±0.07 | 0.28±0.07 | 0.61±0.08 | 0.29±0.07 |
| **330** | 0.56±0.07 | 0.29±0.07 | 0.6±0.08 | 0.29±0.07 |
| **340** | 0.57±0.07 | 0.3±0.07 | 0.59±0.08 | 0.29±0.07 |
| **350** | 0.57±0.08 | 0.3±0.07 | 0.49±0.07 | 0.24±0.06 |
| **360** | 0.57±0.08 | 0.3±0.07 | 0.66±0.09 | 0.33±0.08 |
| **362** | 0.57±0.08 | 0.3±0.07 | 0.49±0.07 | 0.25±0.06 |
| **372** | 0.57±0.08 | 0.31±0.07 | 0.57±0.08 | 0.29±0.07 |

Table 17B: Heritability ($h^2$), permanent ($p^2$) and common ($c^2$) environment effects estimated at different ages from homogenous and heterogeneous variance model (HOM/HET)

| Age | HOM | | HET | |
|---|---|---|---|---|
| | $h^2$ | $p^2$ | $h^2$ | $p^2$ |
| **382** | 0.58±0.08 | 0.31±0.07 | 0.52±0.07 | 0.27±0.06 |
| **392** | 0.58±0.08 | 0.31±0.07 | 0.57±0.08 | 0.29±0.07 |
| **402** | 0.58±0.08 | 0.32±0.07 | 0.61±0.08 | 0.31±0.07 |
| **411** | 0.58±0.08 | 0.32±0.07 | 0.66±0.08 | 0.34±0.08 |
| **421** | 0.58±0.08 | 0.32±0.07 | 0.66±0.08 | 0.34±0.08 |
| **431** | 0.58±0.08 | 0.32±0.07 | 0.63±0.08 | 0.33±0.08 |
| **441** | 0.58±0.08 | 0.33±0.07 | 0.65±0.08 | 0.35±0.08 |
| **451** | 0.58±0.08 | 0.33±0.07 | 0.55±0.08 | 0.29±0.07 |
| **453** | 0.58±0.08 | 0.33±0.07 | 0.59±0.08 | 0.32±0.07 |
| **463** | 0.58±0.08 | 0.33±0.07 | 0.63±0.08 | 0.34±0.08 |
| **473** | 0.58±0.08 | 0.33±0.07 | 0.62±0.08 | 0.34±0.08 |
| **483** | 0.58±0.08 | 0.34±0.07 | 0.58±0.08 | 0.32±0.07 |
| **493** | 0.58±0.08 | 0.34±0.07 | 0.61±0.08 | 0.35±0.08 |
| **502** | 0.58±0.08 | 0.34±0.07 | 0.61±0.08 | 0.35±0.08 |
| **512** | 0.57±0.08 | 0.35±0.07 | 0.59±0.08 | 0.34±0.07 |
| **522** | 0.57±0.08 | 0.35±0.07 | 0.57±0.08 | 0.34±0.07 |
| **532** | 0.57±0.08 | 0.35±0.07 | 0.48±0.07 | 0.29±0.06 |

Table 17C: Heritability ($h^2$), permanent ($p^2$) and common ($c^2$) environment effects estimated at different ages from homogenous and heterogeneous variance model (HOM/HET)

| Age | HOM | | HET | |
|---|---|---|---|---|
| | $h^2$ | $p^2$ | $h^2$ | $p^2$ |
| **542** | 0.56±0.08 | 0.36±0.07 | 0.54±0.08 | 0.34±0.07 |
| **544** | 0.56±0.08 | 0.36±0.07 | 0.59±0.08 | 0.37±0.08 |
| **554** | 0.56±0.08 | 0.36±0.07 | 0.59±0.08 | 0.38±0.08 |
| **564** | 0.55±0.08 | 0.37±0.07 | 0.57±0.08 | 0.38±0.07 |
| **574** | 0.55±0.08 | 0.37±0.07 | 0.59±0.08 | 0.41±0.08 |
| **584** | 0.54±0.08 | 0.38±0.07 | 0.58±0.08 | 0.41±0.08 |
| **593** | 0.54±0.07 | 0.38±0.07 | 0.51±0.07 | 0.38±0.07 |
| **603** | 0.53±0.07 | 0.39±0.07 | 0.52±0.07 | 0.39±0.07 |
| **613** | 0.52±0.07 | 0.4±0.07 | 0.53±0.07 | 0.42±0.07 |
| **623** | 0.52±0.07 | 0.41±0.07 | 0.51±0.07 | 0.42±0.07 |
| **633** | 0.51±0.07 | 0.41±0.07 | 0.52±0.08 | 0.45±0.07 |
| **684** | 0.47±0.07 | 0.46±0.07 | 0.48±0.08 | 0.52±0.08 |
| **694** | 0.46±0.07 | 0.47±0.07 | 0.47±0.07 | 0.53±0.08 |
| **704** | 0.45±0.07 | 0.48±0.07 | 0.46±0.07 | 0.54±0.07 |
| **714** | 0.44±0.07 | 0.49±0.07 | 0.44±0.07 | 0.56±0.07 |
| **724** | 0.44±0.08 | 0.5±0.07 | 0.43±0.08 | 0.57±0.08 |

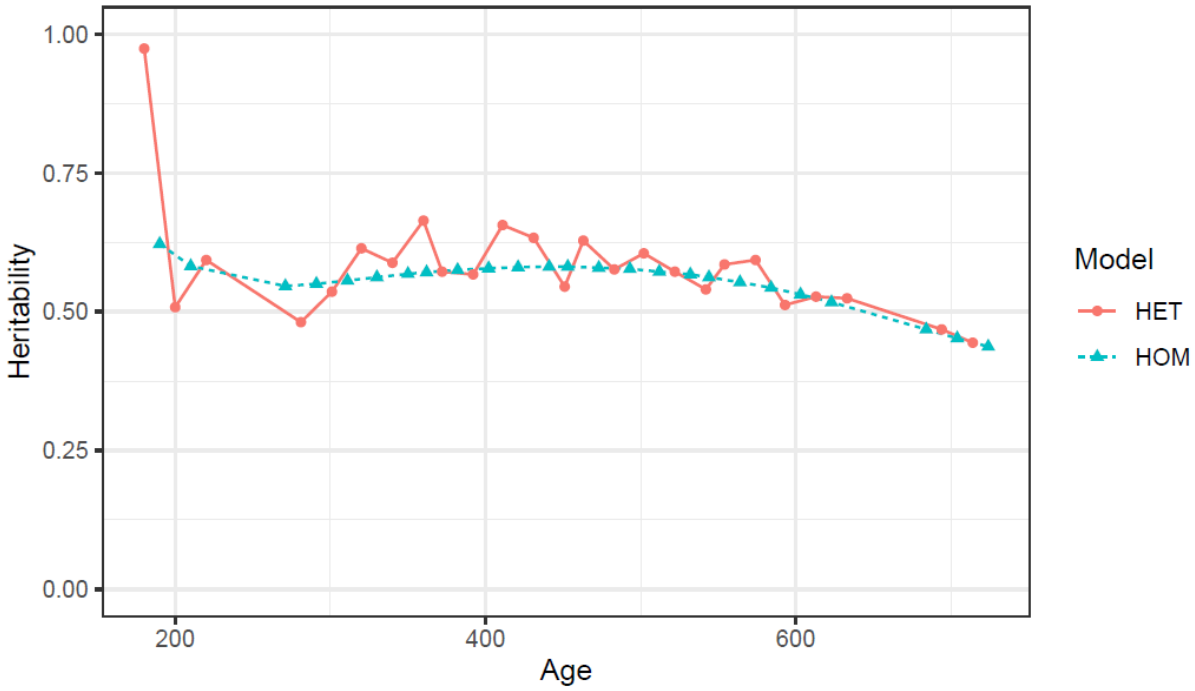



Figure 28: Plot representing the change in heritability over time as estimated from different models. HET and HOM are heterogeneous and homogeneous models

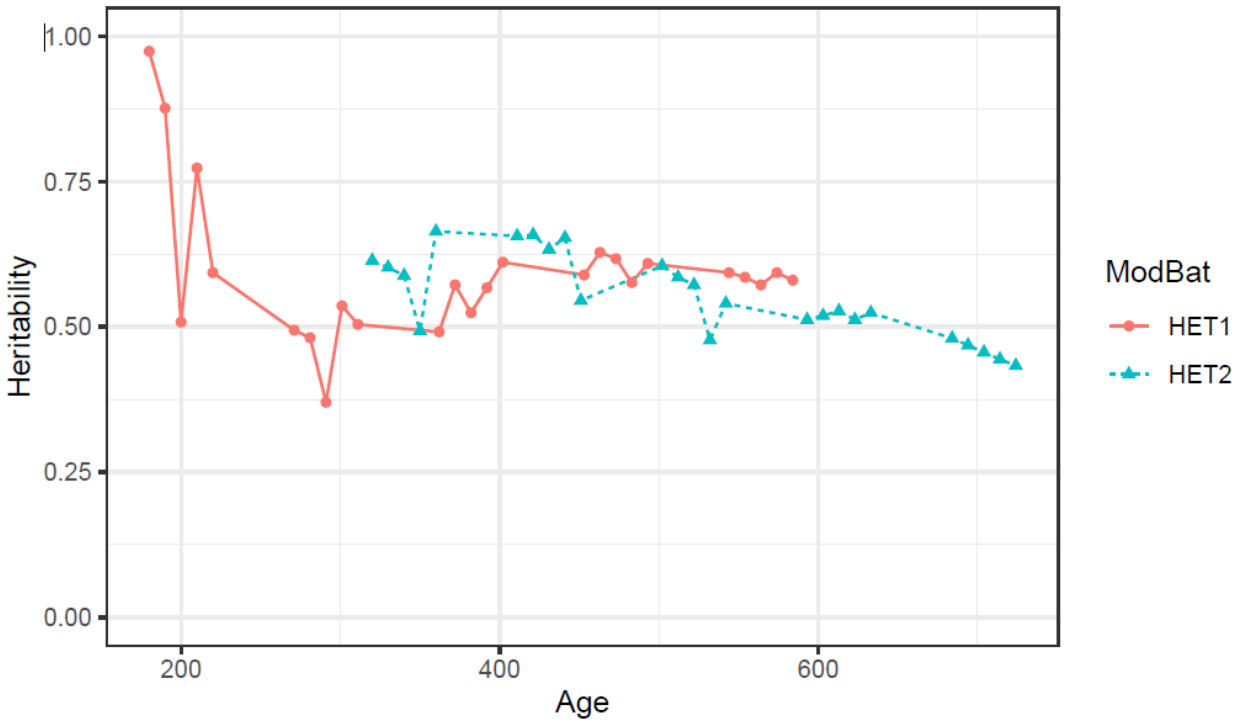



Figure 29: Plot depicting the batch wise change in heritability over time. HET-1 and HET-2 represents the heritabilities estimated from heterogeneous model for batch 1 and 2.

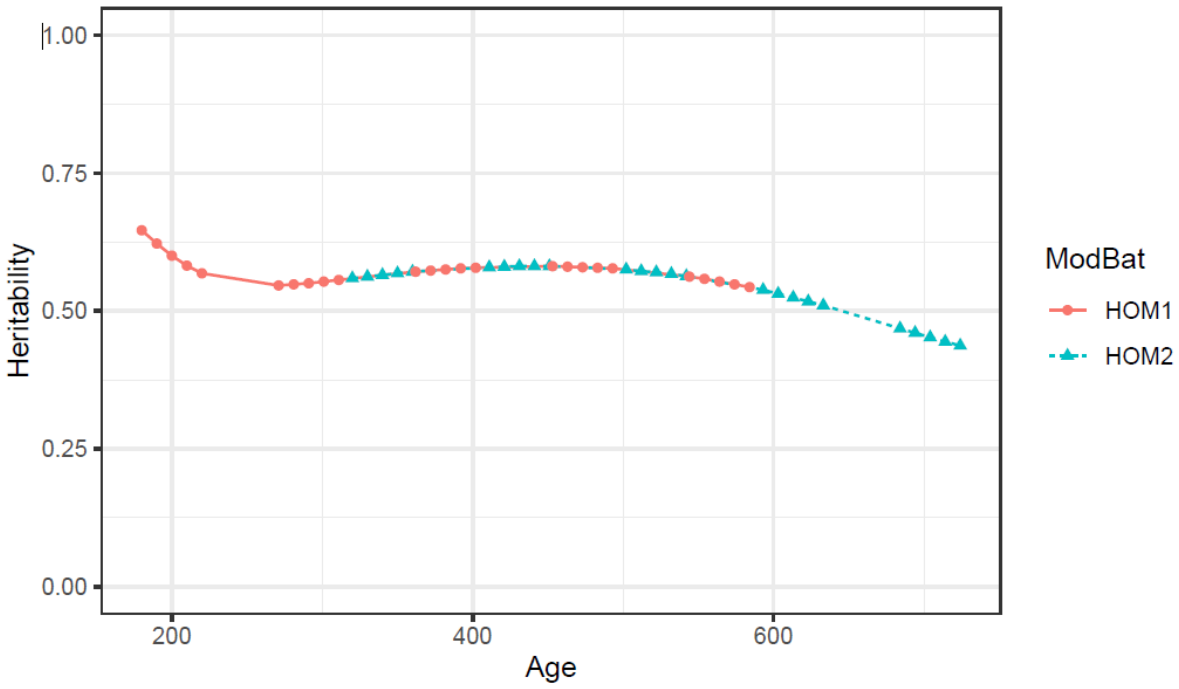



Figure 30: Plot depicting the batch wise change in heritability over time. HOM-1 and HOM-2 represents the heritabilities estimated from homogeneous model for batch 1 and 2.

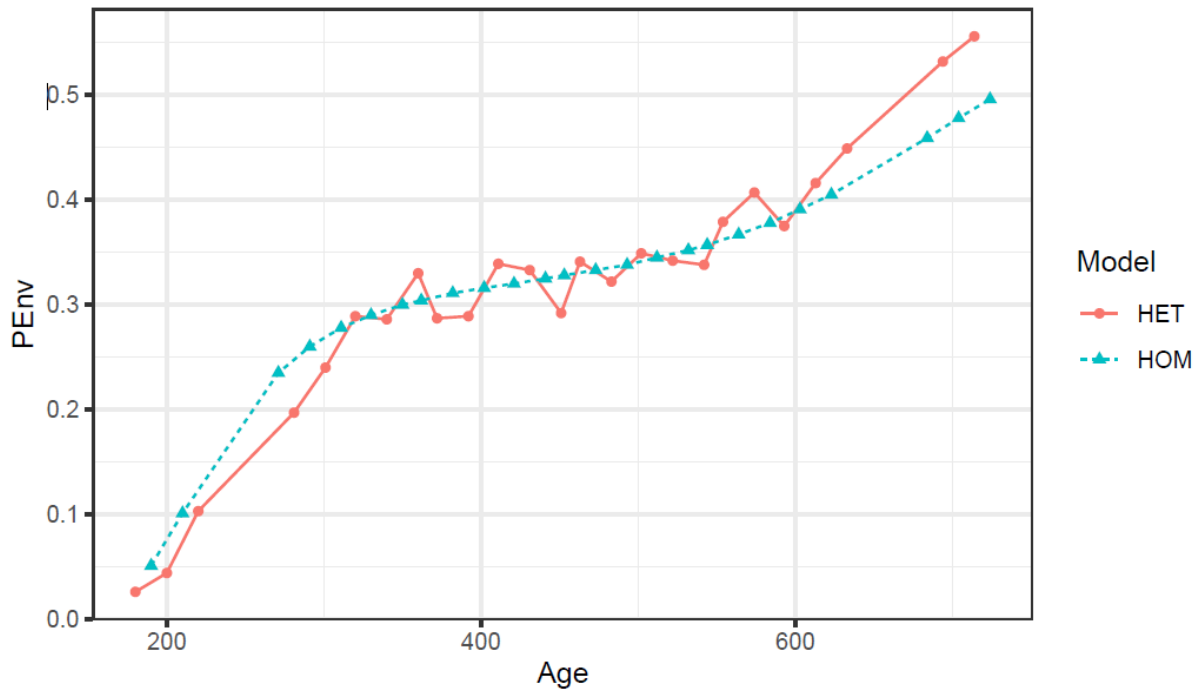



Figure 31: Plot representing the change in permanent environmental (PE) effect over time as estimated from different models. HET and HOM are heterogeneous and homogeneous models

#### 4.8.2.5 Genetic and phenotypic correlations across ages

The correlations between the ratios of genetic, permanent environmental, and phenotypic variances, between 50 values of ages (control variable), estimated from different RR models are illustrated in the contour plots in Figures 32 to 34. For each of the corresponding random effects, the correlation matrix was of the order 50 × 50 with 2500 entries. Hence, it would be more convenient to look at a 3-d plot to have a better understanding of the behavior of correlations. The contour plots depicted in Figures 32 to 34 are 3-d plots collapsed into 2-dimension. Every contour corresponds to an increase or decrease of correlation by 0.1 units marked with a different color gradient. In the contour plots, the diagonal from the top right to the bottom left corner (secondary diagonal) corresponds to a correlation of 1, and it tends to decrease towards either side of the secondary diagonal.

The genetic correlations obtained from the HOM and HET models ranged from -0.1 to 1 respectively (Figure 32). The pattern of change in genetic correlations with age was similar in both the models. The genetic correlations began to decrease between the ages with an increase in the interval between ages. A negative genetic correlation was observed between the first and last ages for both HOM and HET models. In both the cases, the surface of the plots were very steep for the age intervals at the beginning, and at the end, whereas the curvature of the surface was relatively less steep towards the middle of the age classes. A steep curvature corresponds to a sharp gradient, and a lesser steeper surface corresponds to a slow change in correlation values with the change in intervals.

The correlations between permanent environmental effects between different ages varied from -0.3 to 1 for HOM model and 0.4 to 1 for HET model (Figure 33). The change in permanent environmental correlations between different ages was different for HOM and HET models. For HOM model the curvature of the surface was very steep towards either side of the secondary diagonal, whereas, a relatively flatter surface for permanent environmental correlations was observed for HET model.

The phenotypic correlations between different ages as estimated from both the models are presented in Figure 34. The phenotypic correlations ranged from -0.1 to 1 for HOM model and 0 to 1 for HET model. The change in phenotypic correlations were steep towards either side of the secondary diagonals for HOM model, but for HET model, the surface of phenotypic correlations were ridged with irregularities throughout the surface coupled with a steep gradient (Figure 34).

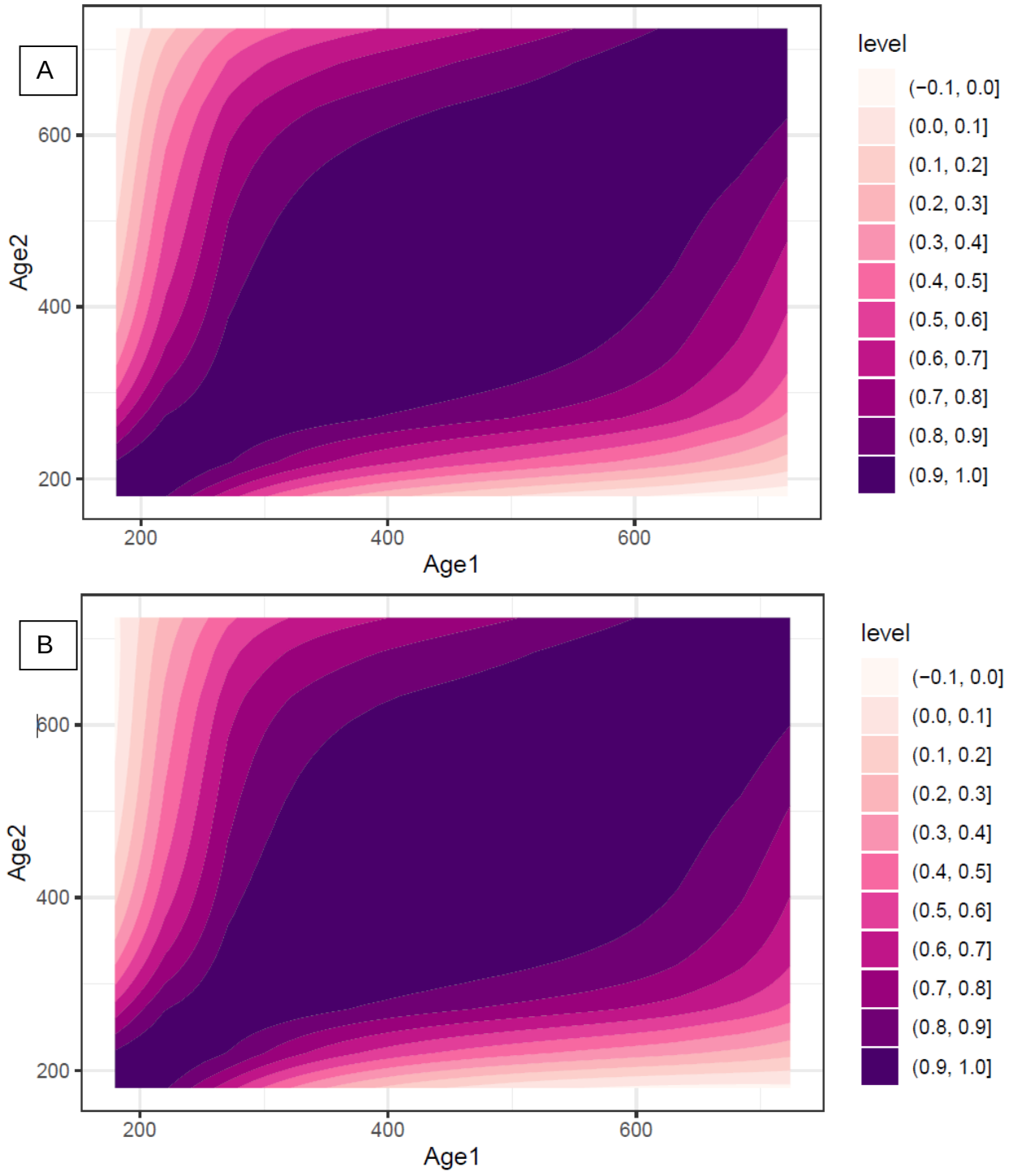



Figure 32: A - Genetic correlations between different ages estimated from homogeneous (HOM) models; B - Genetic correlations between different ages estimated from heterogeneous (HET) models

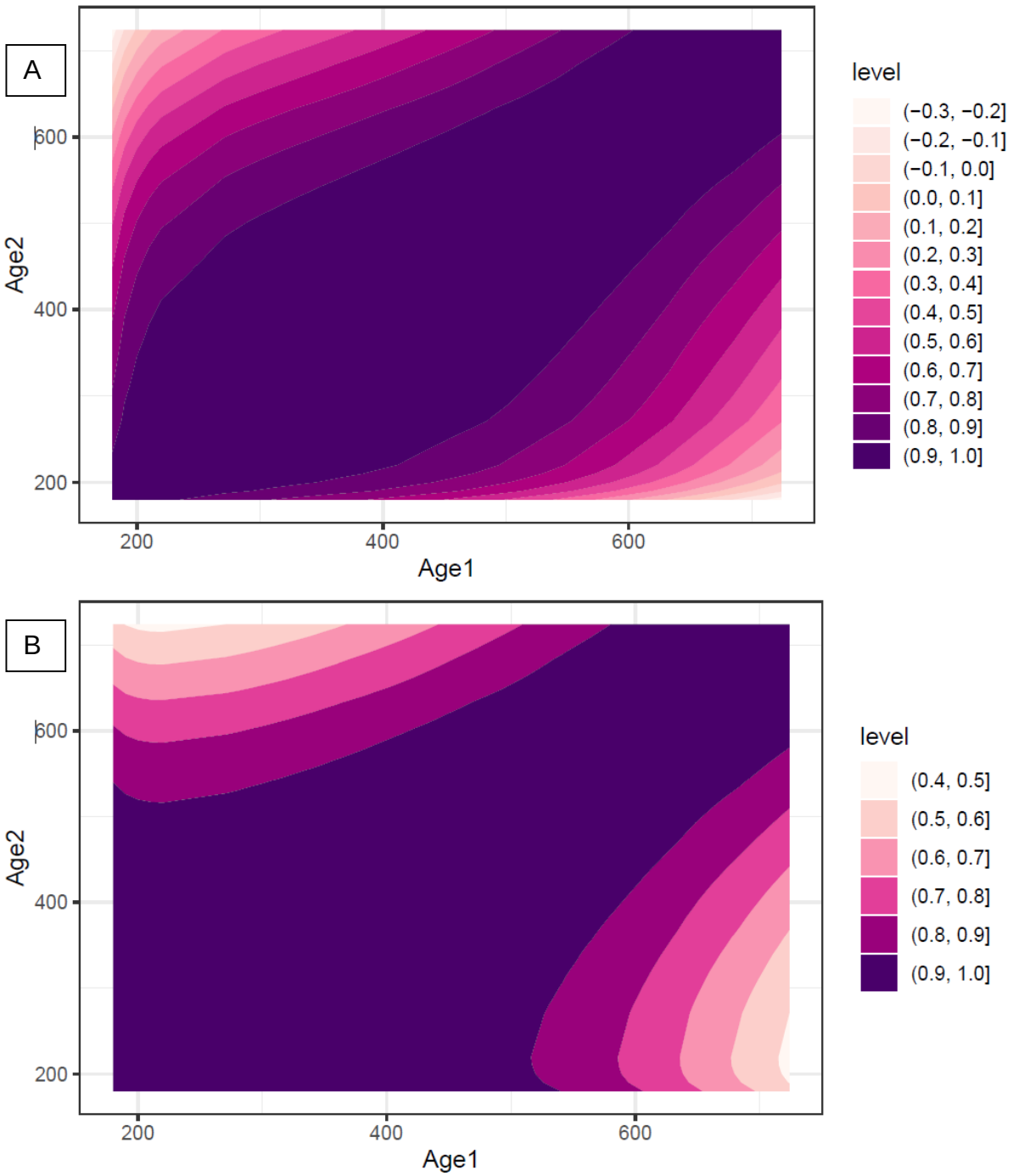



Figure 33: A - Permanent environmental correlations between different ages estimated from homogeneous (HOM) models; B - Permanent environmental correlations between different ages estimated from heterogeneous (HET) models

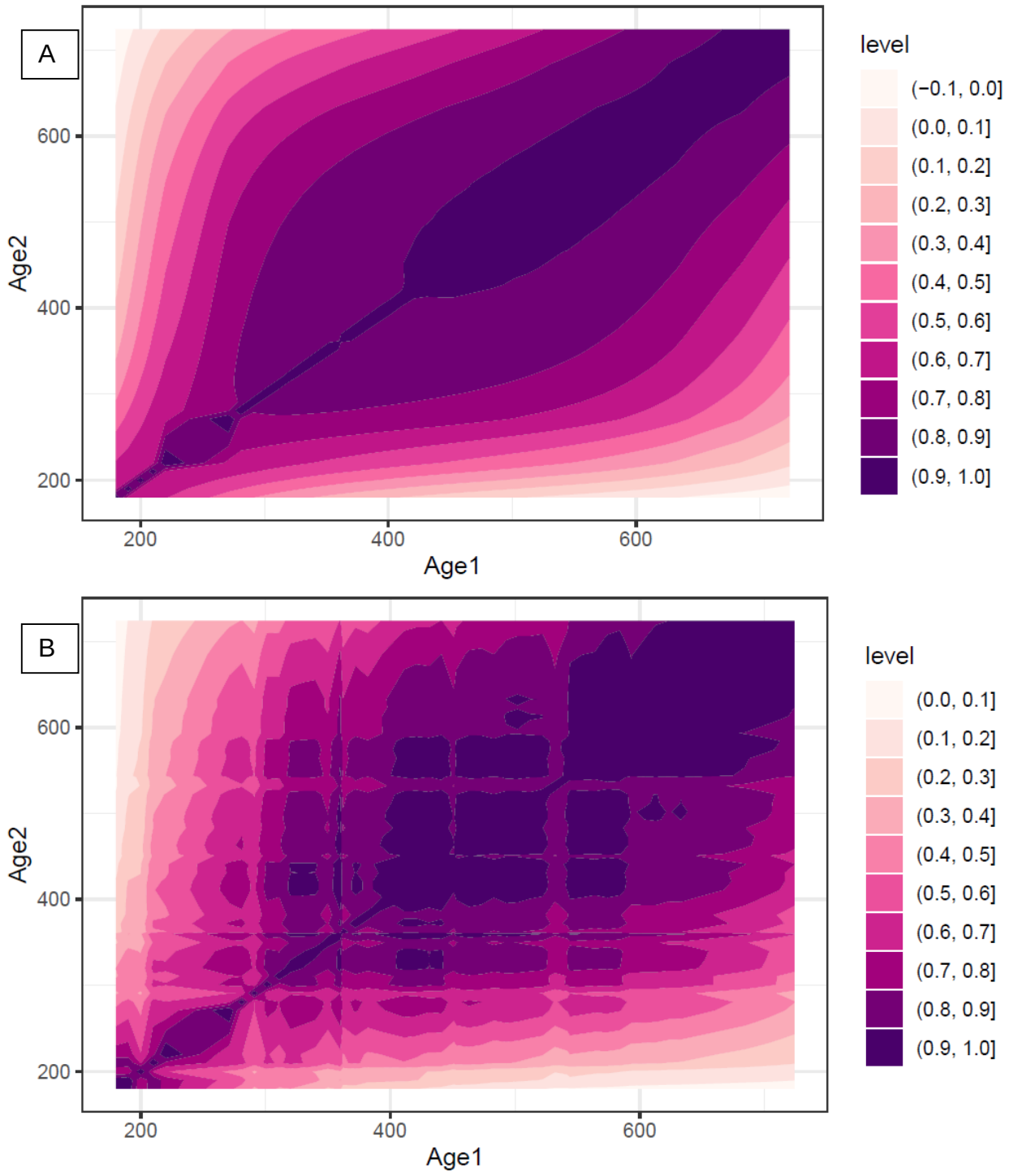



Figure 34: A - Phenotypic correlations between different ages estimated from homogeneous (HOM) models; B - Phenotypic correlations between different ages estimated from heterogeneous (HET) models

#### 4.8.2.6 Eigenfunctions

Three different patterns of variations were modelled by the quadratic random regression model;

(1) The variation in the vertical shift of individual growth trajectories from the mean growth trajectory,

(2) The variation in the slope of individual trajectories form the overall slope of the mean trajectory and

(3) The variation in the curvature of the individual trajectories.

These three patterns of variations were captured and represented as different axes of variations using respective eigenfunctions as depicted in Figure 35. Each eigenfunction is estimated as a linear combination of intercept, slope and curvature. The canonical variate (z1) associated with the largest eigenvalue (820.83 and 791.4) represents the axes of variation corresponding to the factor which are positively correlated at every age. Note that the eigenfunctions associated with the first axes (z1) took positive values at all ages. For the second axes (z2), corresponding to another factor, the eigenfunction took positive values at the earlier age and negative values at the later ages (Figure 35). For the third factor, the eigenfunctions were positive at earlier and later ages, whereas negative at the intermediate ages (Figure 35). The amount of variation associated with each axes was represented by eigenvalues. The majority of the genetic variation in the trajectory could be attributed to the first axis with an associated eigenvalue of 820.83 for the HET model and 791.40 for HOM model. The genetic contribution to the second axes was comparatively less for both the models with the magnitude of eigenvalues of 82.75 and 91.66 for the HET model and the HOM model. The least contribution of the genetic variation was observed for the third axes with the eigenvalues of 25.79 for the HET model and 23.07 for the HOM model.

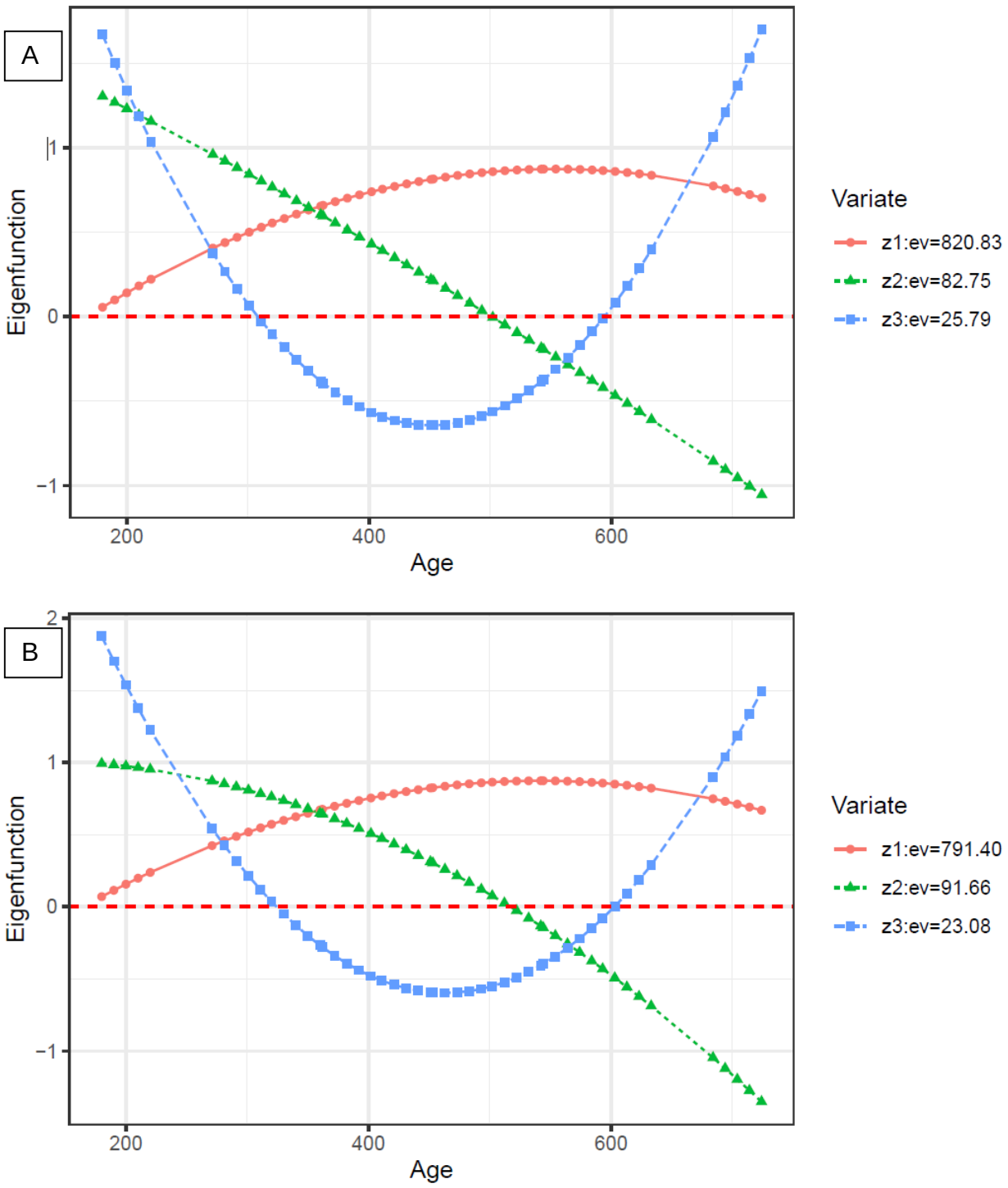



Figure 35: Eigenfunctions based on covariance function of order 3. The lines represent single trait selection response for canonical variates ($\boldsymbol{z_i}$) , each of them with an associated variance (ev = eigenvalue). A - Heterogeneous variance model, and B - Homogeneous variance model.

# 5. DISCUSSION

*Clarias magur*, an air-breathing catfish, is endemic to India (Khan *et al*., 2000) and is well known for its potential for aquaculture. The majority of fish farmers depend on the natural seed of magur for culture. The farmers collect the magur seed from the natural breeding grounds, eventually putting pressure on natural stocks, leading to depletion of natural stocks in many areas (Jousy *et al*., 2018). Also, natural seeds exhibit a wide range of variability in their performance and many times harbor pathogens that are difficult to detect. Controlled breeding is possible in magur, and selective breeding can be implemented to enhance the species' productive and reproductive performance (Jousy *et al*., 2018). The potential of genetic selection to improve the economic traits in the desired direction is well established. It ensures the availability of quality seed on demand, optimizes production efficiency, and most importantly, it reduces the pressure on natural stocks. In any genetic selection program, it is important to quantify the effect of different factors affecting the animal's performance to manage and optimize the production system. Further, it is also important to have the knowledge of genetic parameters of economically important traits since the response upon applying the selection depends on the additive genetic variation present in the population. In the present work, different models and methods were used to evaluate the genetic potential of economic traits in magur to enhance its growth performance by adopting a selective breeding program. First, the effects of various non-genetic and genetic factors on the selected traits of magur at stocking and harvest were quantified. Further, the genetic parameters of different traits at stocking and harvest were estimated using different models. The heritability of body weight at harvest was evaluated with different methods, and the results were compared. The breeding value for body weight at harvest was predicted with both univariate and multivariate models implementing a parental and animal model to compare the accuracies of the predictions. Finally, the body weights recorded at different ages were evaluated with a multi-trait model, and the genetic parameters of growth trajectories were evaluated with the help of random regression models and covariance functions.

Globally, catfish farming is dominated by *Pangasius*, *Clarias,* and *Ictalurus* species. In 2016 the global catfish production was 5.06 million metric tonnes and was dominated by Pangasid catfish followed by Ictalurid and Clarias species (Tacon, 2018). Although global catfish production is significant, the application of genetic selection programs in catfish to improve economically important trait has lagged behind that of species like Atlantic salmon (Houston and Macqueen, 2019), rainbow trout (Sae-Lim *et al.*, 2013), and tilapia (Ponzoni *et al.*, 2011). Several factors resulted in the slow adoption of a pedigree-based breeding program in catfish species. The increase in production of catfish species like *Pangasius* is rapid, mainly due to the focus on the expansion of farming area and operations, improving production techniques, controlling diseases, and developing appropriate feed. Further, the major share of catfish production comes from small and family-owned farms lacking the resources or technical expertise to conduct large-scale breeding programs in contrast to the programs undertaken by large-scale breeding companies for trout, salmon, and tilapia (Bosworth *et al.*, 2020).

However, recently there has been a surge in the publications on genetic improvement programs in various catfish species. A review was published recently on stripped catfish *Pangasiandon hypothalamus*, based on a long term, pedigree-based breeding program by Vu *et al.* (2019). Bosworth *et al*. (2020) reported the heritability and selection response for growth traits and carcass yield for channel catfish. The *Ictalurus punctatus* breeding program was established in 2006 in the United States. Thailand has an on-going genetic improvement program on African catfish *Clarias gariepinus*, and the genetic parameters and genotype-environment interaction for growth traits and the response to selection for growth traits were reported by Srimai *et al.* (2019, 2020). Srisapoome *et al.* (2019) studied the genetic evaluation of disease resistance in bighead catfish, *Clarias macrocephalus*, for which heritability was estimated for immune traits and disease resistance.

### 5.1 Effect of non-genetic factors on growth

The magur fish used in the present study were from two-year classes (2014 and 2015), represented as batch-1 and batch-2, respectively. In both the years, the breeding extended through July- September. As a result, the fish's age and body weight

at tagging varied both within and between the batches, which introduced variation in the initial body weight at stocking due to the age factor. No prior information was available about the survival rate in magur from spawning to age/size at tagging when the present selective breeding program was started. The average size at tagging for batch-1 was 41 g, and for batch-2 was 10 g. To avoid the loss of families, all the spawns from fullsib families were retained and stocked, which led to a significant difference in the stocking density at the initial stages of larval rearing. Factors like differential fecundity, variation in hatching, along with differential mortality across fullsib families, also led to the difference in the number of offspring per family and in turn, the stocking densities at the initial stages of larval rearing. Differential stocking densities resulted in differential growth. The fullsib families with very high stocking density grew slower, and fullsib families with low stocking density grew faster. In the present study, it was observed that the heterogeneity in the body weight at stocking influenced the harvest body weight. The analysis of results showed that the BW0 had a significant effect on BW at harvest ($R^2$= 12.5 %; Table 4); hence, the BW0 was used as the covariate in the model for variance component estimation of traits at harvest. The BW0 was preferred as a covariate over the fish's age because there were only ten levels of age groups compared to 74 levels of BW0, hence the dependency of harvest BW on age at stocking was minimal. Also, the difference in age will get translated into the difference in the BW0, it was decided to use BW0 as the covariate. The BW0 as a covariate substantially reduced the error variance and increased the R-square value of the model in the estimation of the Mean sum of squares. Also, the AIC value for the model containing BW0 was much lower (AIC = 12731.7) compared to the model without BW0 (AIC = 13243.20). The model with the lowest AIC value is considered to give better estimates of variance components (Akaike, 1974). The present study indicated the need to standardize the stocking density of magur during family-wise rearing to minimize the variation in the body weight. Care should be taken to maintain similar stocking density across different families reared in different tanks until they are tagged and stocked for communal rearing.

The genetic selection program for magur was initiated with the objectives to develop magur strains suitable for mono and poly-culture. The fish from batch-1 were reared under both mono and polyculture (with *Labeo rohita*) conditions. In the present

study, it was observed that there was no significant difference in the mean harvest body weight of magur reared in mono or polyculture system (Table 4). Based on the farmers' inputs and considering the low reproductive rate of the species, it was decided to continue genetic selection to develop a strain only for the poly-culture. The non-significant effect of culture type on the harvest body traits is possible because magur is a bottom dweller and competes very little for space with rohu, a column feeder. Jousy *et al.* (2018) had reported a significant difference between the body weight of magur in monoculture and polyculture systems and attributed the difference to the considerable variation in the number of observations between the two culture systems (699 fish under monoculture whereas only 249 fish under poly-culture). In the present study, the number of observations of both culture systems was similar.

The fish from non-AP stock showed higher growth than the fish from AP stock. From Table 2, it is evident that non-AP stock, on an average, has grown 13.46 g heavier than AP stock. It is interesting to note that, AP stock of batch-1 grew 20 g heavier than the AP stock of batch-2. The reasons could be differences in the environment (season, year, management etc.) across two batches, the difference in age at stocking, differences in the genetic makeup of the animals, and also it could be due to the lower weight of the fish at stocking in batch-2. In 2014 when selective breeding was initiated, there were no reports about the size and site for PIT tagging in magur. Hence, it was decided to carry out the experiment to identify the optimum size and appropriate body location for PIT tag insertion in the magur. In half of the magur born in batch-1, the PIT tag was inserted in the muscle, and in another half, it was inserted in the abdominal cavity. To insert the PIT tag in the muscle, the fish need to be large; hence, the tagging was taken up when fish reached an average size of 40 g. The experiment revealed that the magur retained PIT tags inserted in the abdomen reasonably well (over 85%), and tagging can be done in small size fish also, so all the magur belonging to batch-2 were tagged by inserting the PIT tag in the abdominal cavity when the fish were of about 10-15 g size.

Report suggest that magur males grow heavier than females (Jousy *et al*., 2018). In the present population, males were 27 g heavier than females. The difference

in the growth performance between males and females could be attributed to physiological differences. It was observed that the pond environment had a significant effect on body weight at harvest. In aquaculture, pond dynamics is a common phenomenon where each pond has a unique environment due to the variation in the water quality parameters, the difference in pond productivity, pond biota etc., all of which will have an impact on the growth of fish. Further, even though efforts were taken to maintain uniform stocking density across all the ponds, differential mortality altered the uniformity in stoking density over a period. The highest mean weight was observed for the fish from the pond with the lowest stocking density, which was expected. There was a significant difference in the body weight at stocking of the fish of batch-1 and batch-2. At stocking, to ensure that the fish were randomly allocated to different ponds, its effects were tested. It was observed that there was no significant difference between the body weights of magur distributed across various ponds at the time of stocking. Since the ponds were nested within the batch, the comparison of ponds had to be made within batches. Another important reason for the differences in performance may also be attributed to the differences in pond productivity and other factors that contribute to the pond dynamics, which are hardly under human control.

## 5.2 Genetic parameter estimates at stocking and harvest

### 5.2.1 Heritabilities

To initiate the genetic selection program, the formation of a base population with a broad genetic base is essential. To ensure a large genetic base and considering the availability of resources like the ponds for family rearing and lack of information/experience in family-wise breeding of magur, only fullsibs were produced in batches 1 and 2 adopting a single pair mating design. Hence, in the present study, records from only fullsib families were present; as a result, the estimated additive genetic variance is confounded with the variance due to non-additive genetic effects, maternal and common environmental effects (Falconer and Mackay, 1996; Gjedrem, 2005). In an animal model the random effects of interest are the additive genetic value of individual animals (Kruuk, 2004). Two models were used for the estimation of variance

components for traits at harvest. In model-1, only the animal effect was considered as a random effect, and in model-2, along with the animal effect, the pond effect was also considered as a random effect. A note of caution has to be made about the pond effect. In the present study, the ponds used to aquaculture the fish of batch-1 and batch-2 were different and hence were nested within the batch. There was also no link between the families of batch-1 and batch-2, hence there is a possibility that the additive genetic effect may have got confounded with the pond effect, hence it becomes more useful to treat ponds as random to facilitate the recovery of some genetic information if any (Legarra *et al*., 2008). In both, the models the genetic variance and residual variance remained unchanged, which suggests that there was no confounding effect between the pond and additive genetic effects. So from an analysis point of view, the model 1 and 2 may be treated as equivalent.

Heritabilities estimated for traits at stocking were very high, 0.74 ± 0.08 for BW0, and 0.97±0.08 for TL0 and moderate, 0.35±0.05, for K0. Eventually, a decrease in heritability was observed for BW (0.44±0.07), TL (0.32±0.06), and K (0.07±0.02) along with moderate to high heritabilities estimated for BD (0.22±0.04), HW (0.27±0.05), and ADG (0.42±0.07) at harvest (Table 5). The high heritability estimates close to unity at the time of stocking can be majorly attributed to the presence of the high amount of effects common to fullsibs. The fullsib families were reared in separate cement tanks till tagging, for a long time, which might have introduced common environmental effects (confounding of tank effects with additive genetic effects) pertaining to fullsibs, increasing between family variance and thereby inflating the heritability of traits at stocking. The homogeneity of the growth rate within the family could vary depending on the age of the fish, perhaps the growth rate within family was similar at the early stages of life resulting in low within family variance and high heritability, but might have varied as the age progressed (resulting in high within family variance and low heritability). There are various reports regarding the presence or absence of common fullsib effects in aquaculture species. In channel catfish, Reagan *et al*. (1976) reported heritability estimates exceeding unity as they contained large amounts of common environmental variances. The decrease in heritability estimates from stocking to harvest could be attributed to diminishing effects common to fullsibs in the later stage of life (El-Ibiary *et*

*al*., 1978). A very strong effect common to fullsibs for early growth rate but decreasing with time was observed in common carp (Hulata *et al*., 1976). Reports on other aquaculture species like rainbow trout (Herbinger *et al*., 1995), European sea bass (*Dicentrarchus labrax*) (De Leon *et al*., 1998), and black bream (*Acanthopagrus butcheri*) (Doupe and Lymbery, 2005) also support the fact that the effects common to fullsibs occur primarily during the early life stages and tend to dissipate within a few months of growth. Fu *et al*. (2016) reported non-significant common environment effects on growth traits at 10 and 18 months post-hatching. Ponzoni *et al*. (2005) and Nguyen *et al*. (2007; 2010) reported a diminishing common environmental effect in Nile Tilapia with a longer grow-out period. Even though there was a decrease in heritability over time in the present study, it is not sure whether 12 months of communal rearing can completely even out the common fullsib effect.

Further studies are required to quantify effects common to fullsibs. Jousy *et al*. (2018) reported high heritability for the harvest body weight in magur. In the present study, the heritability estimates for harvest BW obtained from both models were high. Moderate to high heritability for harvest trait is very common in fish and has been reported by many authors for different species (Maluwa *et al*., 2006; Vandeputte *et al*., 2008; Nielsen *et al*., 2010; Marjanovic *et al*., 2014; He *et al*., 2015; Li *et al*., 2018). In the present study, the heritability estimate for body weight was less than that obtained by Jousy *et al*. (2018). Usually, for the first few generations, high heritability estimates for body weights are obtained when natural stocks are used to produce families (Jousy *et al*., 2018). In the current study, all the traits were recorded at one year of communal rearing in earthen ponds under commercial conditions. The present estimates may be taken as indicative of the presence of adequate additive genetic variance for the harvest body weight in magur.

Head length and head width are important morphometric traits in catfish because the head weight has a significant contribution (up to 23.6%, El-Ibiary *et al*., 1976) towards the weight at harvest. Head length was only recorded for the 2015 batch fish. A moderate heritability was obtained for head length (0.25±0.08; Table 5). The heritability for head length and percentage head weight in channel catfish was reported

to be moderate to high (El-Ibiary *et al*. 1978). In the present study, the percentage head weight to the total body weight was not calculated because of the limitation to sacrifice fish. The ratio of head length to total length for batch-2 ranged from 19.5 to 30.5%. Body depth can be considered another measure of growth, and in this study, a moderate heritability for body depth was observed. To the best of our knowledge, reports on heritability estimates of body depth in catfish are not available. There are various reports in Nile tilapia (Rutten *et al*., 2005; Fernandes *et al*., 2015) and the GIFT strain of Nile Tilapia (Reis *et al*., 2014; Oliveira *et al*., 2014) where a moderate to high heritability for body depth was estimated. The moderate heritability of body depth and head width, and head length estimated in our study indicates that the trait may respond to selection and may be incorporated in selection objectives.

The heritability of condition factor (K) decreased from stocking (0.35±0.05) to harvest (0.07±0.02). The decreased heritability of K at harvest indicated the decrease in between family variance with age. The residual variance of K remained unchanged at stocking and harvest (Table 3-4) which means that the decrease in heritability at harvest for K was result of decrease in the between family variance. The condition factor is used to compare the well-being of the fish based on the assumption that heavier fish of a given length are in better condition (Froese, 2006). The low additive genetic variation of the condition factor indicated the variation in the well-being of magur in the present study is mostly influenced by non-genetic factors. Barnham and Baxter (2003) noted that the K value of a fish is influenced by age, sex, season, stage of maturation, the fullness of gut, type of food consumed, amount of fat reserve, and degree of muscular development. Hence the heritability results suggested the maintenance of well-being of fish through better management practices. The heritability score for K value obtained in this study was in the range reported for other species (Guan *et al*., 2016; Li *et al*., 2018; Navarro *et al*., 2009; Premachandra *et al*., 2017). The average daily gain (ADG) as a heritable growth-related trait was also investigated in this study, and the heritability estimated for ADG was high (0.42±0.07; Table 5). The high heritability of ADG indicates a good scope to fasten the growth in magur. Li *et al*. (2018) estimated a high heritability (0.46±0.05) for ADG in Japanese flounder, and Quinton *et al*. (2007) estimated moderate heritability (0.26±0.18) for daily weight gain in a salmonid.

### 5.2.2 Genetic and phenotypic correlation

Genetic correlations between traits can occur either due to pleiotropy or gametic phase disequilibrium (Lynch and Walsh, 1998). The former mechanism is a result of complex biochemical, developmental, and regulatory pathways by which a single gene influences multiple traits. The latter mechanism is due to the non-random association of alleles of different loci. The phenotypic correlation arises when the two characters' expression is modified by the same environmental factors operating within individuals (Lynch and Walsh, 1998).

The genetic correlations between the traits at harvest were all positive, and the correlation's magnitude varied between traits (Table 6). A very high positive genetic correlation was observed between body weight and total length at harvest. The very high positive genetic correlation between the body weight and total length indicates that genetic selection on any one trait leads to the improvement of the other, so in the future, the measurement can be restricted to any one of the traits. The high correlation between body weight and total length is conceptually trivial in that the traits are structurally related. In this study body weight showed a high genetic correlation with other body measurement traits. Body weight at stocking and harvest were positive and were highly correlated genetically, which means that in magur, possibly the same set of genes control growth at both early and later stages of life. The fish which are heavier at the time of stocking tends to grow heavier at harvest. In the present study, only a single pair mating design was adopted, and the families obtained were exclusively fullsibs which can bias the genetic parameter estimates. Hence, the present genetic parameters may be considered as indicative only.

### 5.3 Heritability estimates using different methods

The majority of software packages make use of REML for genetic parameter estimation. The REML refers to the restricted part of the likelihood that is not affected/ influenced by fixed effects; hence it is restricted maximum likelihood (Patterson and Thompson, 1971). In REML estimation, the maximization of likelihood can be achieved in different ways depending upon the order of derivatives available (Thompson

and Mantysaari, 1999). The precise class of algorithms are derivative-free (DF) methods, of which simplex/polytope (Nelder and Mead, 1965) and Powell's algorithms (Powell, 1964) are more common. The most popular algorithm involving the first derivatives is Expectation-Maximization (EM REML) algorithm, while the Average-Information algorithm (AI REML) is the most popular involving second derivatives (Misztal, 2008).

The DF approaches are fast for simple models, whereas they are expensive and unreliable for complicated models and have become obsolete (Gilmour *et al*., 1995). The EM algorithm was considered the most reliable but is slow in convergence and does not directly generate the estimate's standard error. The popularity of AI REML is because it is much easier to compute than the alternative second derivative methods like Newton-Raphson and Fisher Scoring and being not much more complicated than EM per iteration and requiring fewer iterations (Gilmour *et al*., 1995). One advantage of using the Newton-Raphson or Average Information algorithm is that the matrix of the second derivatives of the log-likelihood evaluated at the optima known as the Hessian matrix ($\boldsymbol{H}$) is available upon completion of convergence, which is used to estimate reliabilities and to construct confidence intervals of variance components. Serfling (1980) explained with the help of the asymptotic theory of maximum likelihood that matrix $\mathbf{2}\boldsymbol{H}^{-\mathbf{1}}$ is an asymptotic variance-covariance matrix of the estimated parameters of $\boldsymbol{G}$ and $\boldsymbol{R}$. In the breeding experiments, heritability is the primary genetic parameter used to assess genetic improvement potential. Heritability is estimated as a non-linear function (ratio) of variance components (Meyer *et al*., 2013; Edwards, 2017). Under the REML method, the reliability/standard error (SE) of the estimated heritability is obtained from the first-order Taylor series approximation of the function of heritability. Hence, the estimation of reliabilities is based on the large sample (asymptotic) theory, where the test and confidence intervals are based on the asymptotic normality (Serfling, 1980). The estimated standard errors may be unreliable, especially under a small sample size, since these methods are based on approximations (Littell *et al*., 2006). Hence, this study contrasted the variance components and their reliabilities and the precision of heritability estimates obtained from the REML method with the corresponding estimates obtained using various sampling-based methods.

### 5.3.1 Bootstrap estimation

Bootstrap methods are alternative approaches for estimating the reliability/standard error (SE) of parameters and constructing the confidence interval without assuming the normal distribution. Model-based bootstrapping was first developed under simple linear models, which were bootstrap-based on residuals (Efron, 1979; Efron and Tibshirani, 1994). In the simple linear models, the residuals are resampled either from an estimated empirical distribution of the residuals (parametric bootstrap) or can be resampled from the predicted residuals, without any distributional assumptions (non-parametric bootstrap), after model fitting, which is then added onto an estimate of the mean function obtained from the data (Morris, 2002).

In the present study also, the same idea was extended to the animal model by sampling with replacement from predictors of the random effects and residuals for non-parametric bootstrap and from an estimated multivariate normal distribution for parametric bootstrap. In the context of an animal model, the natural choice of predictors for random effects is BLUPs (BLUP bootstrap), which are readily available in software packages and have optimality properties in predicting an individual's random effect. A simulation study by Morris (2002) demonstrated that the optimal properties of the BLUP do not transfer over to bootstrapping when the sample size is small, as a result of which the BLUP-based bootstrap consistently underestimates the variability in the data. Further, he reported that the coverage probability of 90 % intervals for the variance components from BLUP bootstrap showed severe under coverage problems. However, in the present study, based on the results obtained for coverage probabilities (Table 11) there were no under coverage issues with the 95% confidence intervals for variance components and heritability, which add confidence to the optimality of the current sample size (Table 11). An increase in the coverage rate with an increase in sample size was reported by Thai *et al*. (2013). In our study, both non-parametric and parametric bootstrap methods gave similar coverage probabilities (0.98 for non-parametric bootstrap and 0.94 for parametric bootstrap – Table 11) and can be attributed to the total bootstrap sample size (10,000). An increase in the total bootstrap sample size might give exact coverage probability for both non-parametric and parametric bootstrap.

The non-parametric bootstrap estimate of heritability showed an upward bias of 15.9 % compared to the REML estimate (Table 9). Searle *et al.* (2009) explained that the realizations of BLUP are shrinkage estimates and the overall effect of shrinkage estimation is the reduction in the variance surrounding realized BLUPs (Morris 2002). This reduction in variance is translated to every new data set generated by resampling from BLUP predictions. In a mixed model analysis with pedigree, the shrinkage of the BLUP is a function of the number of observations per family, the total number of observations, the observed vector of trait values, and estimated variance components (Searle *et al*., 2009). A possible overall effect when a data set is generated from BLUP-based bootstrap is the reduction in within-family variation, which is evident from the present study results. In the present study (Table 9), there was a major reduction in the residual variance obtained from non-parametric bootstrap than the REML estimate, which resulted in the 15.9 % higher heritability estimated from the non-parametric bootstrap. In contrast, the heritability estimated from the parametric bootstrap was similar to the REML estimate. In parametric bootstrap, the random variables were sampled from a multivariate normal distribution, which resulted in an estimate similar to that of the REML method since the latter method is based on the strong assumption of multivariate normality.

A global non-parametric bootstrap was performed (results not shown), in which the random animal effects and residuals were sampled from the respective vectors of BLUPs and residuals, wherein both the vectors were respectively assumed to be a single sampling unit in contrast to assuming each fullsib family as the basic sampling unit. The global non-parametric bootstrap did not yield any meaningful result, where the estimated variance was near zero. The reason for this is, while performing bootstrapping, one of the concerns is that the resampling should appropriately mimic the true data generating process that produce the true data set (Flachaire, 2005). Hence, it is evident that the classical bootstrap methods developed for simple linear models need to be modified to take into account the characteristics of mixed-effects models (Das *et al.,* 1999). In the setting of an animal model, the within-family correlation structure needs to be taken into account; thereby, each parent and all its offspring will be the appropriate

sampling units for generating meaningful bootstrap samples, rather than an entire vector of animal effects (predicted breeding values) as a whole.

### 5.3.2 Bayesian estimation

Bayesian methods are often touted for their ability to incorporate prior information when available, but the critical utility of these methods is their ability to provide a complete description of the uncertainty of an estimate (Walsh and Lynch, 2018). In the present study, an MCMC chain of 10,00,000 iterations was run with a burn-in of 50,000 iterations (for better convergence and mixing of chain) and a thinning interval of 100 (to reduce autocorrelation), which yielded a total of 9,500 sampled values of variance components and heritabilities. Convergence and autocorrelation are two important issues to monitor when using the MCMC method. An autocorrelation of less than 0.10 in magnitude is considered reasonable (de Villemereuil, 2012). There is a possibility of the strong dependence of the values obtained in the first few iterations on the starting values. The chain is said to converge only after the dependence on the starting parameter has diminished. The visual examination of the time series trace plots suggested a well-mixing chain (Fig 1a-c). As a rule of thumb, if there is no trend in the trace, then the chain has achieved convergence (Hadfield, 2019). Upon convergence, a stationary distribution of chain is generated, and drawing multiple samples from the stationary distribution should yield equal means that can be statistically tested with Geweke's statistic which has an asymptotically standard normal distribution (Plummer *et al*., 2018).

One of the results of the large sample theory is the asymptotic normality of the posterior distribution, i.e., to say as $n \to \infty$ (as more and more data arrive from the underlying process), the posterior distribution of $\theta$ approaches normality (Gelman *et al*., 2013). Examination of the distribution of the variance components (additive and residual; Fig 2a-b) obtained from the MCMC sampler showed that the posterior distribution was approximately normal. The additive variance obtained from the REML analysis was in close agreement with the location parameters of the posterior distribution, whereas the REML residual variance was notably less in comparison with the Bayesian posterior

estimate. The probable reason could be that a weak informative inverse gamma prior was used to obtain the posterior distribution. Gelman (2006) noted the popularity of the inverse gamma family of priors is due to its clean mathematical properties, wherein he discourages the use of inverse gamma prior due to the sensitivity of its parameterization towards the posterior inferences. However, in the present study, the MCMCglmm package of R software used did not offer the flexibility to use a different family of prior other than inverse gamma. Waldmann and Ericsson (2006) reported a high residual variance in the posterior distribution when comparing the REML and Gibbs sampling estimates of genetic parameters. However, according to Sorenson and Gianola (2002), the variance components from REML analysis should be identical to the mode of Bayesian posterior distribution if mixed models' parameters are assigned non-informative uniform distributions. A study in turbot fish Guan *et al*. (2017) showed that the Bayesian estimate of additive variance from the posterior mean was high compared to the REML estimate, whereas the Bayesian residual variance was low in comparison with the corresponding REML estimate. In the present study, a higher additive and residual variance were obtained from the posterior mean, median, and mode compared to the REML estimates (Table 9). In a simulation study, Van *et al*. (1996), obtained similar estimates for the variance components from the posterior mean and REML, and the variance components obtained from the posterior mode were always low. In our present study, despite the difference in residual variance, the mode of the posterior distribution of heritability resulted in similar value as that of the REML estimate, giving more confidence in the estimates. Several other studies also reported where both REML and Bayesian methods gave a similar estimate of heritability if the sample size is large and heritability is high (Waldmann and Ericsson 2006; Alijani *et al*., 2012; De Villemereuil *et al*., 2013; Guan *et al*., 2017). It should be noted that the standard error estimated from the posterior distribution was similar to the one obtained from the delta method, the approximate 95% CI for REML heritability agreed well with the 95% CI of HDR of heritability of posterior distribution. An objective comparison between the estimates of the REML and Bayesian posterior distribution is difficult. However, the heritability estimates obtained from the REML and Bayesian methods were close to each other, which indicates that the likelihood function has an overpowering influence on the prior

distribution. The influence of the prior distribution on the posterior diminishes under a large sample size (Walsh and Lynch, 2018); nevertheless, only a prior sensitivity study can ascertain this claim as noted by Blasco (2017).

### 5.3.3 Asymptotic sampling

Mayer and Houle (2013) described a simple alternative to estimate the sampling distributions of the functions of variance components by repeated sampling of parameter estimates from their asymptotic, multivariate normal (MVN) distribution. Later, the functions of interest (heritability) for each variance component sample can be estimated and inspect for their distribution across replicates. A sampling of REML estimates from asymptotic MVN distribution, specified by the inverse of the information matrix, offers a straightforward and computationally undemanding way to derive sampling distributions and confidence intervals for estimates of covariance components and their functions (Mayer and Houle, 2013). In our study, we sampled 10000 estimates of both additive genetic and residual variances and estimated heritability for the combination of every sampled value. The mean of the variance components and heritability estimates obtained from the asymptotic sampling were similar to the REML estimates of variance components and heritabilities (Table 9-10). The standard error for variance components and heritabilities obtained from both methods were similar. In the present study, both the REML and asymptotic sampling performed equally well to estimate uncertainties in variance components and their functions. An admonition in sampling from asymptotic distribution is that, to yield a valid estimate of covariance, sampling distributions, and confidence intervals, large sample properties should hold well, i.e., the inverse of the information matrix has to provide an adequate description of sampling covariance among the parameters estimated. The inverse of the information matrix could be obtained with Newton-Raphson type algorithms or its variant, especially the average information algorithm (AIREML) (Gilmour *et al*., 1995), which utilizes second derivatives of log-likelihood.

The uncertainties surrounding the REML heritability could be inaccurate when the assumed asymptotic behavior is violated. Schweiger *et al*. (2016) describe

different cases under which the asymptotic normality does not hold well, of which the most common cause is the sample size. They further demonstrated that the asymptotic CIs tend to be biased when the heritability estimated is relatively low or high, under which the CIs can spread beyond the natural boundaries of their parameters (e.g., negative heritabilities). In a breeding program, the number of families and number of offspring per family that could be generated at a point in time is constrained by many factors. As noted previously, under large sample assumptions, it is not clear what constitutes a large sample (Walsh and Lynch, 2018), making it hard to realize the goodness of asymptotic approximation. As an alternative, sampling-based approaches can be used to generate a full distribution of the estimates to construct CIs.

Among different methods used in the present study, the parametric bootstrap and asymptotic sampling gave similar results for variance components and heritability as the REML estimates. Even though the methods were different, the underlying assumptions are the same, hence gave rise to similar results. Non-parametric bootstrap is well-known for estimating the uncertainties in the data by generating a distribution of estimates. In our study, we have used the BLUP predictions for resampling to generate new data sets. However, BLUP is a shrinkage estimator, and as a result, there was a shrinkage in the residual variance resulting in a slightly upward bias in the non-parametric bootstrap estimate of heritability in comparison to the REML estimate of heritability. A possible alternative could be to generate bootstrap samples by resampling the deviations of phenotypic records from the estimated fixed effects and then estimating the bootstrap samples' parameters. Finally, the Bayesian posterior mode yielded a heritability estimate similar to the REML heritability; however, there were slight differences in the residual variances obtained from Bayesian and REML methods. Among the methods used in our study, the results indicated the Bayesian methods to be more robust, followed by non-parametric bootstrap. It is less likely that the parametric bootstrap and asymptotic sampling differ with REML estimates, and hence we recommend the Bayesian method as a robust alternative to REML when the information content of the data is questionable. A more structured way to find the adequacy of data might be, to sensibly use different prior distributions to check its effects on the results. The data with adequate information should overcome the effect of prior such that the

divergence between the prior and posterior distribution is the amount of information contained in the data.

### 5.4 Prediction of breeding values

The breeding values for individuals and families corresponding to four traits recorded at harvest viz BW, TL, BD, HW and two derived traits at harvest viz K, and ADG were predicted with univariate and multivariate BLUP models. In a combined selection method, the selection decision is based on both the family and individual BVs (Falconer and Mackay, 1996). Selection based on multiple information leads to improved genetic gain compared to stand-alone methods. In the present study, the information from both fullsib (parental) and animal models were used to predict the BVs for different traits. Further, an animal model also predicts the BVs of the parents, with or without the records. Since the present data belonged to the base population, there were no phenotypic records available for the parents. In the present study, the accuracy of BV of parents predicted from the animal model was low, probably due to the lack of records for the parents; hence the parental model was used to predict the BV of families.

#### 5.4.1 Comparison of predicted BVs

The breeding values were predicted using both univariate and multivariate BLUP models. Since different traits considered at harvest are recorded from the same individuals, they tend to correlate with each other. Further, the heritabilities were different for different traits at harvest (0.07-0.44), hence it is of interest to understand the differences in breeding value predictions while using a multivariate model over the univariate predictions. The breeding value predictions from either model were compared in terms of accuracy of predictions, prediction error variances, and rank correlations. One of the advantages of multivariate predictions of breeding value is the increased accuracy of predictions (Mrode, 2014). Isik *et al.* (2017) noted that the low heritable traits gains increase in accuracy in a multivariate evaluation by exploiting their correlation with high heritability traits. In the present study, an average increase in estimated BV's accuracy by 13.1% from the parental model and 25.8% from animal model was observed for condition factor (K), which had the lowest estimated heritability (0.07±0.02) when the

multivariate model was employed. The lowest improvement in accuracy (0.36 % and 1.45 % increase from parental model and animal model, respectively) was observed for the body weight, which had the higher observed heritability of 0.44±0.07. The increase in accuracy is directly proportional to the reduction in prediction error variance (PEV). The BLUP methodology predicts the breeding values with minimum variance, with the consequence of BLUP being a shrinkage estimator (Searle *et al.*, 2009). The amount of shrinkage depends on the number of offspring per family and the heritability estimate. For a trait with high heritability, the amount of shrinkage will be higher as compared to a trait with low heritability, and when the shrinkage is high, the PEV is less, and when shrinkage is less, the prediction error will be high. It indicates that when a trait has a high heritability, the scope of improving its accuracy through multivariate evaluation is less; however, when the trait has a low heritability, the multivariate evaluation gathers information from traits with high heritability and increases the accuracy.

In the present study, there was a negative association between the number of offspring per family and the absolute difference between the BV predicted from univariate and multivariate BLUP models. As the number of offspring per family increased, there was less difference between the breeding values evaluated from univariate and multivariate models. Adding more traits into the model apparently increased the number of observations per family (family size), resulting in less difference in the predicted breeding values estimated from univariate and multivariate models. The BLUP breeding values tend to shrink more as the family size increases, thereby reducing the PEV and increasing the accuracy of prediction. The BLUP in the context of an animal model can be interpreted as the linear regression of phenotype on the pedigree (de los Campos *et al*., 2009). However, the BLUP is not a robust regression method since the linear regression method under Gaussian assumption is sensitive to outliers (Hampel *et al*., 2011; Lange *et al*., 1989; Seber and Lee, 2012). The data outlying in case of animal breeding can be genetic or environmental in origin. Even though the significant outliers were removed ad-hoc post analysis, there is always a chance for concealed substructure or underlying heterogeneity, e.g. competition for food or social interactions, which is not accounted for in the experimental design. Such an undeclared preferential treatments can lead to data outlying and thereby deviations from multivariate normality. As a result

of such deviations resulting from random factors, the superimposing of numerator relationship matrix on phenotypes, would result in the shrinkage of performance records to the family means (Gianola *et al*., 2018). The level of shrinkage is again a function of family size and heritability (Searle *et al.*, 2009). In the present study, the unbalanced design and a varying range of trait values across fullsib families led to differential shrinkage of BVs among families when evaluated using univariate and multivariate models. As a result, no clear pattern of change in the ranks of families was observed when evaluated with univariate and multivariate models. The results of rank correlations between univariate and multivariate BVs were very high (0.99) for BW for which the heritability was 0.44±0.07, whereas the lowest rank correlation was 0.80 for the trait K with a heritability of 0.07±0.02. The high and positive values of rank correlations between univariate and multivariate BVs indicate that there is no significant change in the ranking of BVs between univariate and multivariate models.

Mrode (2014) noted that the gain in accuracy is proportional to the absolute difference between the genetic and residual correlations between the traits. In the current investigation, the trait condition factor (K) achieved the maximum increase in accuracy, followed by the body depth (BD). The absolute difference between the genetic and residual correlation was the highest for K (difference of 0.86 – Table 12B) followed by BD (difference of 0.52 – Table 12B). For a trait with a low heritability score, Schaeffer (1984) explained that, while the absolute value of error correlation is less than the absolute value of genetic correlation then the trait with lower heritability achieves the greater percentage reduction of PEV and vice versa. In this work, the absolute value of error correlation was always less than the genetic correlation (Table 12C) between pairs of traits (except between K and TL), and the gain in accuracy was consistently high for traits with low heritability. However, for the pair total length (TL) and K, even though the absolute value of error correlation was higher (0.59, Table12C) than genetic correlation (0.25, Table 12C), the percentage reduction in PEV was more for the trait K, which had a low heritability than for TL which had high heritability. This is in contrast with the explanation of Schaeffer (1984) as mentioned previously, since when anbsolute error is higher than absolute genetic correlation the trait with high heritability should gain more accuracy. However, in the current analysis, six traits were evaluated simultaneously, so

the K had always a chance to pair with the trait where the absolute value of error correlation, which was lower than the absolute value of genetic correlation, hence it attained the maximum accuracy. Out of the five different combinations the K had its absolute value of genetic correlation higher than their corresponding absolute error correlations in four cases. The results suggest that when more than two traits are recorded, it is desirable to analyze all traits together, rather than uni or bivariate analysis, for the breeding value predictions.

Meyer (1983) examined data structure's effect in improving prediction accuracy, i.e., through more connections between the fixed effects and sires. In her work, it was shown that an improvement was achieved in the data structure (more connections between data and more observations per family) by including multi-traits is as important as extra genetic information gathered from multiple traits in increasing the accuracy of prediction. In our analysis, the multivariate model improved the data structure through more connections between fixed effects and families. Further, non-zero residual covariance between the traits resulting from the use of an unstructured covariance matrix also improved data connections. Further, in this study, it was noticed that an increase in accuracy was always more when an animal model was used in comparison to parental models. In the study reported by Schaeffer (1984) observed that the effect of error and genetic correlations are more prominent with animal model (owing to the numerator relationship matrix) than the parental model; hence the accuracy of genetic evaluation may be more in an animal model than the parental models. In the present study, the multivariate prediction of BV involving different traits showed an improvement in the accuracy, especially a substantial increase in prediction accuracies for the traits K and BD, which were having low heritability compared to other traits. An increase in prediction accuracy will result in a higher genetic gain, hence, it is desirable to implement a multivariate model for the evaluation of breeding values when records on different traits are available with different heritability scores.

## 5.5 Genetic parameter estimates of BW at different ages

A multi-trait model was used to genetically evaluate the body weights recorded on the same individuals at five different ages. The heritability at all five different ages was high, and it decreased from stocking to harvest (0.84 to 0.62, Table 13). The corresponding heritability estimated from single-trait models also followed the same trend (0.91-0.64) and not very different from multi-trait heritability estimates (Table 13). The genetic correlations between all body weight at different ages were positive and high (0.56 – 1.00). A perfect hundred percent genetic correlation was observed between BW4 and BW5, which indicates that the selection can also be applied at nine months to improve the harvest body weight. The present study suggests that there is scope to minimize the age of selection for improving the harvest body weight. The reduction in the culture period of the test population will help to reduce the cost of the breeding program and increase the genetic gain by reducing the generation interval. The phenotypic correlations between body weight measurements at different ages ranged from 0.49 to 0.95. This implies that within the first three months of communal rearing there was a notable decrease in the high amount of variation in the initial body weights at stocking. The permanent environmental effect increased with the age of fish (0.07 to 0.19), however from six months to twelve months of culture period, there was no increase in the permanent environmental effects (0.18 – 0.19).

## 5.6 Genetic evaluation of growth trajectories

The growth trajectory is another important economic trait and can be used as a selection objective if adequate additive genetic variation exists (Schaeffer, 2016). The repeated measurements of body weight from the same individual at different ages give rise to the growth trajectory (Meyer and Hill, 1997). Though repeated measurements are assumed to be under the repeated influence of fixed parameters, they are also under the influence of random additive genetic effects, random permanent environmental and residual effects leading to variation in growth trajectories among animals (Van der Werf and Schaeffer, 1997). In contrast to selecting for growth at a fixed time for improving the overall performance of the population, the selection for growth trajectory can result in a

correlated response for improved growth at infinitely many ages, and the growth between any selected time can be optimized (Van der Werf, 2002)

In the present study, two different univariate random regression models (RRM) were used to estimate the heritabilities of body weight measurements recorded repeatedly over the culture period and genetic correlations between pairwise bodyweights at specific days of age. The quadratic Legendre polynomials were used to model various random effects about the growth trajectories. Both homogenous (HOM) and heterogeneous (HET) residual variance models were implemented for genetic evaluation of growth trajectories. The heterogeneous variance model was found to be the best model based on the Bayesian information criteria (BIC) (Table 16). Based on the results of the RRM, it was found that the heritabilities tend to decrease with age, which is at odds with the general norm of decreasing heritabilities with the progression of age as reported earlier (McKay *et al.*, 2002; Rutten *et al.*, 2005; Turra *et al.*, 2012; He *et al.*, 2017; Zhao *et al.*, 2018). One reason for this trend in the present study could be the presence of high common environmental effects resulting from the separate rearing of families in cement tanks for a long time. In this study, there was no replication given at the time of separate rearing of families and hence was not possible to account for the same, which inflated the heritability estimated at the early age.

The estimated heritability for body weight at the time of harvest from the univariate animal model was the same as that of the heritability obtained from RRM model beyond 700 days of age (Table 16). There was a decreasing trend in heritability even towards the final ages, which is indicative either of an increasing within family variation or a decreasing between family variations or increasing permanent environment effect towards the harvest or a combination of these. In this study, the residual variance from the heterogeneous variance model exhibited a decrease in the residual variance towards the final ages (Figure 26). However, the additive genetic variance showed an increasing trend up to 500 days of age and a decreasing trend beyond 550 days of age (Figure 23). This rules out the possibility of an increase in within-family variation as a reason for the decrease in heritability at later ages, which also indicates that a notable amount of common environment effects have been evened out

as the age advanced. All animals have a biological maximum for growth determined by genetics. As the age advances, it allows the individuals to make up their slow growth and reach the genetic potential, thus reducing the variation in population; this may be one reason for the decline in the heritability as age advances. The decrease in heritability towards final ages may also be due to an increase in permanent environmental effect coupled with a slight decrease in the additive genetic effect.

Schaeffer (2016) was of the opinion that looking at the heritabilities of the parameters of the random regression coefficient is probably the best way to look into RRM rather than just estimating the heritabilities at different ages. In the present study, from the results of the percentage contribution of genetic Eigen values (Table 16), the majority of genetic variation observed among growth trajectories while using HOM model (87.34 % of genetic variation) and HET model (88.32 % of genetic variation) was pertaining to the parameter intercept of RRM (Table 16). This means that there is a scope to improve the growth in magur by vertically shifting the mean growth trajectory through selection on the intercept. The percentage contribution of genetic Eigenvalues to slope and curvature showed that the slope and curvature's genetic variation was 10.12% and 2.5 % for the HOM model, 8.9 % and 2.77 % for HET models (Table 16), respectively, and were comparatively lesser than the intercept. This implies that along with vertically shifting the mean growth trajectory, there is a further possibility to improve the growth through increasing the slope (growth rate) of the mean growth trajectory. The positive genetic correlation between the intercept and slope, 0.54±0.09 for HOM and 0.59±0.08 for HET models (Table 16), proposes the scope for simultaneous selection for improvement of both intercept and slope of growth trajectories. Negative genetic correlation of the curvature of the trajectory between both the intercept (HOM: -0.72; HET: -0.70) and slope (HOM: -0.27; HET: -0.43) suggests that a selection applied on either intercept or slope will decrease the curvature of the growth trajectories. The curvature of the trajectory represents the non-linearity in growth rate, hence selecting for intercept or slope will produce the animals with a more linear growth rate. The results of permanent environmental effects suggest that 90 % of the permanent environmental effect was pertaining to the intercept and ≈ 10 % to the slope of trajectory, while there was no permanent environmental effect on the curvature of trajectory. This entails that,

in this study, a linear Legendre polynomial was sufficient to model the permanent environmental effects of the observed growth trajectories. In the present study, both genetic and permanent environmental effects were modeled with the same order (quadratic) polynomial to avoid any computational issues with regard to model convergence.

A decreasing pattern of genetic correlations, estimated from HOM and HET models, were observed between pairwise bodyweights as the interval between specific days of age increased. The decrease in the genetic correlations was rapid between the initial ages up to 300 days, beyond which the rate of decrease in genetic correlations between ages dropped quickly. Thus, the genetic selection for bodyweight at an age less than 300 days may not improve the harvest body weight as desired. Also, beyond 600 days, the genetic correlations between the adjacent ages began to decrease more rapidly than between 300 to 600 days. The heritability and genetic correlations between adjacent ages were higher between 400 to 600 days (Figure 32). Accordingly, a selection applied at 400 days of age or beyond can maximize the correlated response at later ages (up to 600 days of age). The results suggest an early age selection may be considered for faster improvement of the growth in magur and reduce the breeding program's cost.

### 5.6.1 Eigenfunction analysis

The growth trajectories can increase the economic value of domesticated species by altering growth patterns through artificial selection (Kirkpatrick *et al.*, 1990). Kirkpatrick and Heckman (1989) proposed that the coefficient of the covariance matrix (Table 16) can be used to analyze the patterns of inheritance by decomposing the coefficient matrix into its respective eigenfunctions and eigenvalues. The eigenfunction represents a possible evolutionary deformation of the mean growth trajectory, which is paired with respective eigenvalues that are proportional to the amount of genetic variation in the population corresponding to its eigenfunction. In the present study, the eigenfunctions and corresponding eigenvalues of the covariance functions estimated from both HOM and HET models were evaluated (Figure 35). It should be noticed that the sign of the evaluated values between eigenfunctions is irrelevant (for example, the

first eigenfunction in this study has only positive values), but what matters is how the values of the eigenfunctions change over the trajectory (the second eigenfunction from this study changes from positive to negative values along trajectory) (Van der Werf and Schaeffer, 1997). In this study, the first eigenfunction is the most probable deformation of the mean growth trajectory when subjected to selection, which involves an overall increase or decrease of size at all ages. The large size of the first eigenvalue indicates that selection will produce rapid changes if this kind of alteration in the mean growth trajectory is favored. Further, it also means that the major part of the genetic variance is explained by a factor (equivalent to principal component – a linear combintion of intercept, slope and curvature effects) that is constant for all ages, and selection on this factor will increase weight at all ages. The shape of the first eigenfunction also suggests a decrease in growth beyond 600 days. The second eigenfunction corresponds to genetic changes that increase (or decrease) size up to 500 days of age, and decrease (or increase) beyond 500 days. The third eigenfunction showed a more complex pattern. However, the second and third eigenvalues reveal that the amount of genetic variation associated with these eigenfunctions is small compared to the variation associated with the first eigenfunction. The different eigenvalues indicate that, the evolutionary response to selection would be many orders of magnitude faster for the selection involving the first eigenfunction than that involving the second and third eigenfunctions.

In magur, and more generally in aquaculture, the selection objective for genetic improvement of growth can either to increase the overall growth by the end of the culture period or to attain the preferred market size as early as possible. The former objective can be achieved by selecting for growth at a specific point, whereas the latter objective can be realized by selecting on the growth trajectories. Selection for body weight at a fixed age will certainly improve the average body weight of the population, provided the presence of additive genetic variation, but it need not result in the same response for the increase in the body weight at an early age. The preferred market size for magur is 150 to 200g, and hence increasing mean body weight beyond 200 g may not be preferable. Hence, the selection objective can be to attain the preferable market size as early as possible. The results from RRM indicate that there is a scope to improve the growth rate to obtain higher body weight in lesser time by selecting the growth

trajectories. Hence, based on the RRM study, it would be more desirable to select for growth trajectories rather than applying selection for growth at any particular fixed time point.

## 5.7 Conclusions

Indian aquaculture is mainly consist of low-value, high-volume three Indian major carps and common carp production. There is a need for the diversification of fish species for aquaculture. Catfish species are the choice for diversification as they have the potential for high growth and can be cultured under different aquaculture practices. In the recent past, the aquaculture of pangasius is coming up on a large scale in India. However, it has its own problems, and also, it is an exotic species. *Claria magur*, popularly known as magur, has many economic and biological features, making it a potential candidate species for aquaculture of all forms viz., mono and poly culture, intensive, semi-intensive, and extensive culture. The species can be marketed at the varying size and in live conditions easily. The species is distributed pan India, and distinct geographical stocks are available with broad genetic variation. Magur matures under pond conditions, and controlled breeding is possible. At present magur aquaculture is not on large scale due to the non-availability of the quality seed in the desired quantity. Genetic selection can improve this species' overall performance and help Indian aquaculture with one more species for its potential use.

Knowledge of reliable genetic parameters is the key to the success of genetic selection programs. Information on non-genetic factors influencing the economic traits helps to standardize the aquaculture practice. Estimation of unbiased genetic parameters and the ranking of animals based on BVs determines genetic progress aimed. Determining the estimated genetic parameters' accuracy and reliability is challenging, and especially it becomes more challenging when the mating designs have inherent limitations. Identifying the suitable model and method to estimate the accurate genetic parameters is an essential aspect of the breeding programs.

In this work, the genetic potential of magur was evaluated by adopting different statistical models and methods. The effects of non-genetic factors like stock,

batch, culture type, pond, and sex were quantified on the growth of magur. The stock, batch, pond, and sex had a significant effect on the traits studied.

The Non-AP stocks performed better at harvest than AP stock. Formation of the base population with high-performing stocks will be equivalent to several generations of genetic selection and help achieve the selection objective quickly. However, care should be taken to have a broad genetic base in the base population for sustained genetic improvement. The inclusion of more animals from non-AP stock in the breeding nucleus may be considered for a higher growth rate.

There was no significant difference in the growth of magur between mono and poly culture systems. Magur breeding is yet to be standardized, and the production of seed on a large scale is not possible with the present technology. Hence monoculture of magur on a commercial scale may be challenging in the present scenario. However, magur adopts well under the polyculture with the carp species and can be used for culture with the Indian Major Carps. The non-significant effect of culture effects suggests that a single strain may perform better in both the culture types.

In the present study, the weight at stocking had a significant effect (12.5%) on weight at harvest. The long breeding season (July-September), varying stocking density in the family-wise culture, and the differential mortality all led to the significant difference in the body weight at tagging and subsequent stocking. Knowledge of the survival rate of magur in the cement tanks especially till they attain a taggable size will help standardize the stoking density and thereby reducing the variation in the body weight at tagging.

The significant effect of non-genetic factors on growth traits indicates the need for standardizing the management practices. It also suggests the need to identify them and include them in the statistical model for estimating unbiased genetic parameters.

Estimation of accurate heritability is the basic requirement of the genetic selection program. Heritability estimate is influenced by several factors viz., the family

structure, sample size, model, and the methods employed for estimation. The heritability estimates for the traits at harvest were moderate to high except for the condition factor. In the present study, only fullsib families were produced, due to which the complete separation of additive genetic effect from other genetic effects was not possible, and it may have biased the heritability on the upper side.

Variance components and their functions are essential parameters of interest in genetic selection. Linear mixed model (LMM) analysis has become a popular tool for analyzing breeding data. The LMM analysis involves the estimation of variance components and the prediction of breeding values assuming the estimated variance components are correct, and the REML is the method of choice to estimate the variance components in pedigreed selection experiments. Often, there is a lack of sufficient data to fulfil the conditions of optimal likelihood estimates. Therefore, the use of alternative approaches for the evaluation of uncertainties in variance components and their functions is desirable. Sampling-based methods are an option for estimating uncertainties associated with the parameter estimates. It was hypothesized that if the sample size does contain sufficient information, the estimates obtained from various approaches should essentially agree. The heritability of body weight at harvest estimated by ANOVA, REML, parametric bootstrap, asymptotic sampling, and jackknife method was similar (0.43-0.45). The non-parametric BLUP-based bootstrap gave a slightly higher heritability estimate as compared to other methods due to the shrinkage of residual variance. The similarity of heritability score gives confidence in the sample size's adequacy used to conduct the present work. The results indicated the Bayesian methods to be more robust, followed by non-parametric bootstrap. It is recommended that the Bayesian and non-parametric bootstrap methods may be employed as alternative to REML when the information content of the data is questionable.

Multivariate evaluation is considered as the optimum methodology to evaluate the animals' genetic potential as it increases the accuracy of evaluations. The multivariate models are especially useful for the traits with low heritability as the accuracy of the estimates for these traits improves substantially. The improvement in the accuracy in multivariate predictions results from the reduction in the prediction error variance due to an improved data structure. Usually in the breeding program several traits are measured and it will be important to evaluate all the traits simultaneously since these traits may be genetically correlated. In the present study the accuracy of breeding values predicted from multivariate model were high for the traits (K and BD) with low to moderate heritability compared to the accuracy of breeding values predicted from the univariate models. The rank correlations between the breeding values predicted from univariate and multivariate models for respective traits were positive and high for both animal model (0.80 to 0.98) and parental model (0.86 to 0.99). In the present study the heritability of the growth traits were high, also the genetic correlation between them was high. Hence the multivariate analysis did not add significantly to the accuracies of univariate estimates. However, it is always desirable to implement a multivariate prediction model when different traits are recorded on the same individual, and especially when the heritabilities are different.

The selection objective in the present breeding program is to improve the body weight at harvest. The discussion with the farmers and other stakeholders suggest that a magur with a body weight more than 200 g is not preferred by the Indian consumer as it is confused with the African catfish or its hybrid. However the opinions differ. This poses a challenge to set the appropriate selection objective. If the BVs can be estimated for the animals for the body weight at various age then the animals can be selected at any stage. Selection for growth trajectory may help in this aspect. Hence in the present study a random regression model was used to estimate the genetic parameters for the growth trajectories. The analysis of results of RRM implied a very high likelihood to increase the growth in magur by changing the growth trajectory through genetic selection.

The present study noted that magur has the qualities to be an important aquaculture species in India and has genetic potential to implement the selective breeding program for the improvement of growth traits at various age group. The study found that the REML method estimates reliable genetic parameters even under small sample size. However, the adequacy of the estimated genetic parameters from a small sample size may be confirmed by using Bayesian and/or non-parametric bootstrap methods to repose the confidence in the estimated genetic parameters.

# 6. SUMMARY

*Clarias magur* is well adapted to a wide range of climatic conditions. It matures under pond conditions, and controlled breeding is possible, which provides an opportunity to improve the overall performance of magur through a genetic selection program. However, regardless of the potential of magur to be a candidate species for all forms of aquaculture, the limited availability of seeds and low quality preventing the large-scale aquaculture of magur in India. In aquaculture, genetic selection is a proven means to ensure the availability of quality seeds in optimal quantities and sustainable aquaculture. A thoroughly optimized genetic selection program can enhance current magur production in India by multiple folds. To maximize selection response from a genetic breeding program, the knowledge of reliable genetic parameters of important economic traits is crucial. Hence the present work aimed to study different non-genetic factors affecting the growth traits in magur and evaluate the genetic potential of magur by employing different models and methods of estimation. Different non-genetic factors considered in the study were stock, batch, culture type, pond, sex, and stocking BW as a linear covariate for the traits at harvest. Various growth traits as in body weight (BW), total length (TL), body depth (BD), head width (HW), condition factor (K), and average daily gain (ADG) were genetically evaluated at stocking and harvest. Also, BW measurements repeatedly recorded at five different time points were assessed with single-trait and multi-trait models. Further, the growth trajectories of magur were also considered economic traits and evaluated their genetic potential using different random regression models.

Among non-genetic factors, batch and sex significantly affected the growth traits at stocking, whereas harvest traits were significantly affected by batch, stock, pond, and sex. The difference in body weight at stocking contributed to a significant variation (12.5 %) to the body weight at harvest. The present genetic selection program of magur was aimed to produce strains suitable for both mono and polyculture systems. The results indicated no significant difference in the harvest body weight of magur cultured in both the systems; hence, a single strain can perform equally well in both the culture

systems. The significant difference in growth across batches could be attributed to the difference in the environments between year classes. The non-AP stock showed better performance at harvest; hence it might be prudent to include more brooders from high-performing stocks in the nucleus. Further, the significant difference in growth between ponds revealed very high dynamics across pond environments. Also, it was observed that on an average, the males tend to grow heavier than the females by 27.31 g.

The heritability estimates for traits at stocking was very high for BW0 (0.74), TL0 (0.97), and moderate for K0 (0.35), which decreased towards harvest for BW (0.44), TL (0.32), BD (0.22), HW (0.27), K (0.07) and ADG (0.42). The very high heritabilities at stocking were due to the high common environmental effects resulting from the separate rearing of families until tagging, which eventually diminished with time due to long communal rearing. The high heritabilities at harvest indicated the presence of additive genetic variance in the base population of magur. Notwithstanding, the heritability estimated in this work is towards the upper bound of parameter space mainly because of the single pair mating design adopted. The genetic correlations between growth traits were positive and high (0.80 to 0.99) so as phenotypic correlations (0.34 to 76).

The heritability of harvest BW was estimated using different methods viz., ANOVA, REML, parametric and non-parametric bootstrap, asymptotic sampling jackknife, and Bayesian methods. Among different methods, ANOVA (0.45±0.08), REML (0.44±0.07), parametric bootstrap (0.44±0.07), asymptotic sampling (0.44±0.07), jackknife (0.43±0.01), and Bayesian posterior (0.43±0.07) gave similar results, whereas non-parametric BLUP based bootstrap the heritability was high (0.51±0.05). The similitude of heritability obtained from different methods imparts more credence in the amount of information in the present study's data. The high heritability from non-parametric BLUP-based bootstrap was based on shrinkage in residual variance emanated from the bootstrap samples generated using BLUP predictions. Among the methods used in our study, the results indicated the Bayesian methods to be more robust, followed by non-parametric bootstrap.

The accuracy of breeding values predicted from multivariate model were high for the traits with low to moderate heritability compared to the accuracy of breeding values predicted from univariate models. The improved accuracy in multivariate predictions resulted from the reduction in the prediction error variance due to an improved data structure. The rank correlations between the breeding values predicted from univariate and multivariate models for respective traits were positive and high for both animal model (0.80 to 0.98) and parental model (0.86 to 0.99), wherein parental models were more stable across univariate and multivariate models. It is always desirable to implement a multivariate prediction model if different traits are recorded on the same individual, especially when the heritabilities are different. Also, it can improve the data structure through non-zero error correlations to reduce the prediction error.

The multi-trait model implemented for repeated records of body weights from stocking to harvest indicated a unit genetic correlation between nine months body weight and harvest body weight evincing prospects of early age selection, which can curtail the cost incurred in the rearing of tagged animals. A random regression model with a heterogeneous variance structure gave the best fit for growth trajectories of magur. For the genetic evaluation of growth trajectories, a quadratic Legendre polynomial was used within the random regression model to estimate trajectory parameters. The estimated parameters upon plugging into the covariance function resulted in the genetic parameter estimated at fifty different ages in magur. The resulting heritabilities tend to decrease with age due to high common environmental effects at the initial ages. The genetic correlations between adjacent ages were less below three hundred days, whereas the genetic correlations were high between four hundred to six hundred days, suggesting the selection within the range will result in a maximum correlated response. The analysis of eigenfunctions implied a very high likelihood to increase the growth in magur by changing the mean growth trajectory through genetic selection. The present study revealed a high amount of additive genetic variance in the base population of magur, the results of which can be utilized to optimize the current breeding program to produce a fast-growing high-performance strain of magur through genetic selection.

The study suggests that the sampling-based methods can be effectively used to understand whether the sample size is large enough to contain sufficient information under likelihood estimation assumptions. The results from the different sampling-based methods indicate that the present sample size is adequate to estimate a reliable additive genetic variance and the heritability estimated is reliable. The present study found that the heritability for the body weight in magur is moderately high, and it can be improved further by genetic selection. The Multivariate analysis of different traits at harvest has improved prediction accuracy of breeding values, hence this approach can produce high response to selection. The RRM models proved the genetic potential of magur to attain higher growth in lesser time and the positive correlated response for higher growth at different ages.

# 7. REFERENCES


Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control, 19*, 716-723.

Alijani, S., Jasouri, M., Pirany, N., & Kia, H. D. (2012). Estimation of variance components for some production traits of Iranian Holstein dairy cattle using Bayesian and AI-REML methods. Pakistan Veterinary Journal, 32(4), 562-566.

Barnham, C. A., & Baxter, A. F. (2003). *Condition factor, K, for salmonid fish*: Department of Primary Industries.

Becker, W. A. (1992). Manual of quantitative genetics. 5th Edition. Academic enterprises, USA.

Blasco, A. (2001). The Bayesian controversy in animal breeding. *Journal of Animal Science, 79*(8), 2023-2046.

Blasco, A. (2017). *Bayesian data analysis for animal scientists*: Springer.

Blasco, A., Sorensen, D., & Bidanel, J. P. (1998). Bayesian inference of genetic parameters and selection response for litter size components in pigs. *Genetics, 149*(1), 301-306.

Borah, B. C. (2020). Asian Catfish Clarias magur (Ham), a Wonder Fish for Health and Nutrition. *Acta Scientific Nutritional Health, 4.2*, 139-143.

Bosworth, B., Waldbieser, G., Garcia, A., Tsuruta, S., & Lourenco, D. (2020). Heritability and response to selection for carcass weight and growth in the Delta Select strain of channel catfish, Ictalurus punctatus. *Aquaculture, 515*, 734507.

Casella, G., & Berger, R. (1990). Statistical Inference. Wadsworth and Brooks/Cole. *Statistics/Probability Series. Pacific Grove, California*.

Chowdhary, S., & Srivastava, P. (2013). Length–weight relationships (LWR) of threatened Asian catfish, Clarias batrachus under poor availability in natural conditions from Unnao, Uttar Pradesh, India. *Advances in Applied Science Research, 4*(6), 138-114.

Copas, J. B. (1983). Regression, prediction and shrinkage. *Journal of the Royal Statistical Society: Series B (Methodological), 45*(3), 311-335.

Corbeil, R. R., & Searle, S. R. (1976). Restricted maximum likelihood (REML) estimation

of variance components in the mixed model. *Technometrics, 18*(1), 31-38.

Das, S., & Krishen, A. (1999). Some bootstrap methods in nonlinear mixed-effect models. *Journal of Statistical Planning and Inference, 75*(2), 237-245.

De Boor, C. (1980). *A practical guide to splines* (Vol. 27): springer-verlag New York.

De Leon, F. G., Canonne, M., Quillet, E., Bonhomme, F., & Chatain, B. (1998). The application of microsatellite markers to breeding programmes in the sea bass, Dicentrarchus labrax. *Aquaculture, 159*(3-4), 303-316.

De los Campos, G., Gianola, D., & Rosa, G. J. (2009). Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation. *Journal of Animal Science, 87*(6), 1883-1887.

De Villemereuil, P. (2012). Estimation of a biological trait heritability using the animal model. *How to use the MCMCglmm R package*.

De Villemereuil, P., Gimenez, O., & Doligez, B. (2013). Comparing parent–offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation study for Gaussian and binary traits. *Methods in Ecology and Evolution, 4*(3), 260-275.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological), 39*(1), 1-22.

Doupe, R. G., & Lymbery, A. J. (2005). Additive genetic and other sources of variation in growth traits of juvenile black bream Acanthopagrus butcheri. *Aquaculture Research, 36*(7), 621-626.

Edwards, S. M. (2017). Standard errors of heritability estimates. Retrieved from http://www.iysik.com/dmu/heritabilities

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. Annals of Statistics, 7, 1-26.

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics* (pp. 569-593): Springer.

Efron, B., & Tibshirani, R. (1994). An Introduction to the Bootstrap (Chapman & Hall/CRC Monographs on Statistics & Applied Probability): Chapman and Hall/CRC.

Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics,*

*3*(1), 1-21.

El Halimi, R. (2009). *Nonlinear mixed-effects models and bootstrap resampling: analysis of non-normal repeated measures in biostatistical practice*: VDM Publishing.

El-Ibiary, H. M., Andrews, J. W., Joyce, J. A., Page, J. W., & Deloach, H. L. (1976). Sources of variations in body size traits, dress-out weight, and lipid content and their correlations in channel catfish, Ictalurus punctatus. *Transactions of the American Fisheries Society, 105*(2), 267-272.

El-Ibiary, H. M., & Joyce, J. A. (1978). Heritability of body size traits, dressing weight and lipid content in channel catfish. *Journal of Animal Science, 47*(1), 82-88.

Elvingson, P., & Johansson, K. (1993). Genetic and environmental components of variation in body traits of rainbow trout (Oncorhynchus mykiss) in relation to age. *Aquaculture, 118*(3-4), 191-204.

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics*: Pearson Education India.

Farias, T., César, J., & Silva, L. (2017). Methods of Selection Using the Quantitative Genetics in Aquaculture-A Short Review. *Insights Aquacult. Biotech, 1*, 1-8.

Fernandes, A. F. A., de Almeida Silva, M., de Alvarenga, E. R., de Alencar Teixeira, E., da Silva Junior, A. F., de Oliveira Alves, G. F., . . . Turra, E. M. (2015). Morphometric traits as selection criteria for carcass yield and body weight in Nile tilapia (Oreochromis niloticus L.) at five ages. *Aquaculture, 446*, 303-309.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 222*(594-604), 309-368.

Flachaire, E. (2005). Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap. *Computational Statistics & Data Analysis, 49*(2), 361-376.

Foulley, J.-L. (2015). Mixed Model Methodology. Part I: Linear Mixed Models. 191. https://www.researchgate.net/publication/271372856_Mixed_Model_Methodology_Part_I_Linear_Mixed_Models doi:10.13140/2.1.3072.0320

Foulley, J. L. (1993). A simple argument showing how to derive restricted maximum likelihood. *Journal of Dairy Science, 76*(8), 2320-2324.

Froese, R. (2006). Cube law, condition factor and weight–length relationships: history,

meta-analysis and recommendations. *Journal of applied ichthyology, 22*(4), 241-253.

Fu, J., Shen, Y., Xu, X., & Li, J. (2016). Genetic parameter estimates for growth of grass carp, Ctenopharyngodon idella, at 10 and 18 months of age. *Aquaculture, 450*, 342-348.

Gallardo, J. A., Lhorente, J. P., & Neira, R. (2010). The consequences of including non-additive effects on the genetic evaluation of harvest body weight in Coho salmon (Oncorhynchus kisutch). *Genetics Selection Evolution, 42*(1), 1-8.

García-Celdrán, M., Ramis, G., Manchado, M., Estevez, A., Afonso, J., & Armero, E. (2015). Estimates of heritabilities and genetic correlations of carcass quality traits in a reared gilthead sea bream (Sparus aurata L.) population sourced from three broodstocks along the Spanish coasts. *Aquaculture, 446*, 175-180.

Gelman, A. (2005). Analysis of variance—why it is more important than ever. *The annals of statistics, 33*(1), 1-53.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis, 1*(3), 515-534.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*: CRC press.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. *Bayesian statistics, 4*, 641-649.

Gianola, D., Cecchinato, A., Naya, H., & Schön, C.-C. (2018). Prediction of complex traits: robust alternatives to best linear unbiased prediction. *Frontiers in genetics, 9*, 195.

Gianola, D., & Fernando, R. L. (1986). Bayesian methods in animal breeding theory. *Journal of Animal Science, 63*(1), 217-244.

Gianola, D., Rodriguez-Zas, S., & Shook, G. (1994). The Gibbs sampler in the animal model: A primer. *Séminaire Modele Animal. INRA, La Colle sur Loup*, 47-56.

Gianola, D., & Rosa, G. J. (2015). One hundred years of statistical developments in animal breeding. *Annu. Rev. Anim. Biosci., 3*(1), 19-56.

Gilmour, A. R., Thompson, R., & Cullis, B. R. (1995). Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models.

*Biometrics*, 1440-1450.

Gjedrem, T. (1983). Genetic variation in quantitative traits and selective breeding in fish and shellfish. *Aquaculture, 33*(1-4), 51-72.

Gjedrem, T. (2012). Genetic improvement for the development of efficient global aquaculture: a personal opinion review. *Aquaculture, 344*, 12-22.

Gjedrem, T., & AKVAFORSK, Å. (2005). *Selection and breeding programs in aquaculture* (Vol. 2005): Springer.

Gjedrem, T., & Baranski, M. (2010). *Selective breeding in aquaculture: an introduction* (Vol. 10): Springer Science & Business Media.

Gjedrem, T., & Robinson, N. (2014). Advances by selective breeding for aquatic species: a review. *Agricultural Sciences, 5*(12), 1152.

Gjerde, B., Simianer, H., & Refstie, T. (1994). Estimates of genetic and phenotypic parameters for body weight, growth rate and sexual maturity in Atlantic salmon. *Livestock Production Science, 38*(2), 133-143.

Goddard, M. (1986). Discussion summary Part VI: Selection and non-random mating. *Advances in statistical methods for genetic improvement of lifestock. Springer, Berlin*, 474-475.

Graser, H.-U., Smith, S., & Tier, B. (1987). A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *Journal of Animal Science, 64*(5), 1362-1370.

Green, B. F., & Tukey, J. W. (1960). Complex analyses of variance: general problems. *Psychometrika, 25*(2), 127-152.

Guan, J., Hu, Y., Wang, M., Wang, W., Kong, J., & Luan, S. (2016). Estimating genetic parameters and genotype-by-environment interactions in body traits of turbot in two different rearing environments. *Aquaculture, 450*, 321-327.

Guan, J., Wang, W., Hu, Y., Wang, M., Tian, T., & Kong, J. (2017). Estimation of genetic parameters for growth trait of turbot using Bayesian and REML approaches. *Acta Oceanologica Sinica, 36*(6), 47-51.

Hadfield, J., Hadfield, M. J., & SystemRequirements, C. (2019). Package 'MCMCglmm'. *R package version*.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (2011). *Robust*

*statistics: the approach based on influence functions* (Vol. 196): John Wiley & Sons.

Hartley, H. O., & Rao, J. N. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika, 54*(1-2), 93-108.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association, 72*(358), 320-338.

Harville, D. A., & Callanan, T. (1990). Computational aspects of likelihood-based inference for variance components. In *Advances in statistical methods for genetic improvement of livestock* (pp. 136-176): Springer.

Harville, D. A., & Carriquiry, A. L. (1992). Classical and Bayesian prediction as applied to an unbalanced mixed linear model. *Biometrics*, 987-1003.

Hayes, J., & Hill, W. G. (1980). A reparameterization of a genetic selection index to locate its sampling properties. *Biometrics*, 237-248.

He, J., Gao, H., Xu, P., & Yang, R. (2015). Genetic parameters for different growth scales in GIFT strain of Nile tilapia (Oreochromis niloticus). *Journal of Animal Breeding and Genetics, 132*(6), 467-474.

He, J., Zhao, Y., Zhao, J., Gao, J., Han, D., Xu, P., & Yang, R. (2017). Multivariate random regression analysis for body weight and main morphological traits in genetically improved farmed tilapia (Oreochromis niloticus). *Genetics Selection Evolution, 49*(1), 80.

He, J., Zhao, Y., Zhao, J., Gao, J., Xu, P., & Yang, R. (2018). Random regression analysis for body weights and main morphological traits in genetically improved farmed tilapia (Oreochromis niloticus). *Journal of applied genetics, 59*(1), 99-107.

Henderson, C. R. (1949). Estimation of changes in herd environment. *J. Dairy Sci, 32*(8), 706-706.

Henderson, C. R. (1950). *Estimation of genetic parameters.* Paper presented at the Biometrics.

Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics, 9*(2), 226-252.

Henderson, C. R. (1963). Selection index and expected genetic advance. *Statisitical*

*Genetics and Plant Breeding*.

Henderson, C. R. (1973). Sire evaluation and genetic trends. *Journal of Animal Science, 1973*(Symposium), 10-41.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 423-447.

Henderson, C. R. (1982). Analysis of covariance in the mixed model: higher-level, nonhomogeneous, and random regressions. *Biometrics*, 623-640.

Henderson, C. R. (1985). Best linear unbiased prediction of nonadditive genetic merits in non-inbred populations. *Journal of Animal Science, 60*(1), 111-117.

Henderson, C. R. (1990). Statistical methods in animal improvement: historical overview. In *Advances in statistical methods for genetic improvement of livestock* (pp. 2-14): Springer.

Henderson, C. R., Kempthorne, O., Searle, S. R., & Von Krosigk, C. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics, 15*(2), 192-218.

Henderson, C. R., & Quaas, R. (1976). Multiple trait evaluation using relatives' records. *Journal of Animal Science, 43*(6), 1188-1197.

Herbinger, C. M., Doyle, R. W., Pitman, E. R., Paquet, D., Mesa, K. A., Morris, D. B., . . . Cook, D. (1995). DNA fingerprint based analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. *Aquaculture, 137*(1-4), 245-256.

Hill, J. R., & Rosenberger, J. (1985). Methods for Combining Data from Gemrplasm Evaluation Trials 1. *Crop Science, 25*(3), 467-470.

Hofer, A. (1998). Variance component estimation in animal breeding: a review. *Journal of Animal Breeding and Genetics, 115*(1-6), 247-265.

Houston, R., & Macqueen, D. (2019). Atlantic salmon (Salmo salar L.) genetics in the 21st century: taking leaps forward in aquaculture and biological understanding. *Animal genetics, 50*(1), 3-14.

Hulata, G., Moav, R., & Wohlfarth, G. (1976). The effects of maternal age, relative hatching time and density of stocking on growth rate of fry in the European and Chinese races of the common carp. *Journal of Fish Biology, 9*(6), 499-513.

Isik, F. (2009). Analysis of half-sib progeny tests in forestry. *Raleigh: North Carolina State University and Departament of Forestry and Enviromental Resources*.

Isik, F., Holland, J., & Maltecca, C. (2017). *Genetic data analysis for plant and animal breeding* (Vol. 400): Springer.

James, W., & Stein, C. (1992). Estimation with quadratic loss. In *Breakthroughs in statistics* (pp. 443-460): Springer.

Jamrozik, J., Bohmanova, J., & Schaeffer, L. (2010). Selection of locations of knots for linear splines in random regression test-day models. *Journal of Animal Breeding and Genetics, 127*(2), 87-92.

Jousy, N., Jahageerdar, S., Prasad, J. K., Babu, P. G., & Krishna, G. (2018). Body weight at harvest and its heritability estimate in Clarias magur (Hamilton, 1822) reared under mono and polyculture systems. *Indian Journal of Fisheries, 65*(2), 82-88.

Khan, M. R., Mollah, M. A., & Ahmed, G. U. (2000). Mass production of hybrid magur and its culture potential in Bangladesh. *Aquaculture Research, 31*(5), 467-472.

Kirkpatrick, M., & Heckman, N. (1989). A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *Journal of mathematical biology, 27*(4), 429-450.

Kirkpatrick, M., Lofsvold, D., & Bulmer, M. (1990). Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics, 124*(4), 979-993.

Kreft, I. G., & De Leeuw, J. (1998). *Introducing multilevel modeling*: Sage.

Kruuk, L. E. (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359*(1446), 873-890.

Kumar, A., Jitender, K., & Hemendra, K. (2017). Length-weight relationship and relative condition factor of Clarius batrachus (Linnaeus, 1758) from Gaurmati Fish Farm, Kawardha, Chhattisgarh, India. *International Journal of Current Microbiology and Applied Sciences, 6*(12), 1425-1431.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 963-974.

LaMotte, L. R. (2014). Fixed-, random-, and mixed-effects models. *Wiley StatsRef: Statistics Reference Online*.

Lange, K. L., Little, R. J., & Taylor, J. M. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association, 84*(408), 881-896.

Lee, C. (2000). Methods and techniques for variance component estimation in animal breeding-Review. *Asian-Australasian Journal of Animal Sciences, 13*(3), 413-422.

Legarra, A., Robert-Granié, C., Manfredi, E., & Elsen, J.-M. (2008). Performance of genomic selection in mice. *Genetics, 180*(1), 611-618.

Li, Y., Zhang, B., Lu, S., Tian, Y., Yang, Y., & Chen, S. (2018). Genetic parameters estimates for growth performance traits at harvest in Japanese flounder (Paralichthys olivaceus). *Aquaculture, 489*, 56-61.

Li, Y., Zhang, B., Yang, Y., & Chen, S. (2019). Estimationof genetic parameters for juvenile growth performance traits in oliveflounder (Paralichthys olivaceus). *Aquaculture and Fisheries, 4*(2), 48-52.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Oliver, S. (2006). *SAS for mixed models*: SAS publishing.

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits* (Vol. 1): Sinauer Sunderland, MA.

Maluwa, A. O., Gjerde, B., & Ponzoni, R. W. (2006). Genetic parameters and genotype by environment interaction for body weight of Oreochromis shiranus. *Aquaculture, 259*(1-4), 47-55.

Marjanovic, J., Mulder, H., Khaw, H., & Bijma, P. (2014). *Genetic parameters for within-family variance of harvest weight in Nile tilapia (Oreochromis niloticus).* Paper presented at the 10th World Congress on Genetics Applied to Livestock Production WCGALP 2014, Vancouver, Canada.

McIntosh, A. (2016). The Jackknife estimation method. *arXiv preprint arXiv:1606.00497*.

McKay, L., Schaeffer, L., & McMillan, I. (2002). *Analysis of growth curves in rainbow trout using random regression.* Paper presented at the Proceedings of 7th World congress on genetics applied to livestock production.

Meyer, K. (1983). Scope for evaluating dairy sires using first and second lactation records. *Livestock Production Science, 10*(6), 531-553.

Meyer, K. (1989). Restricted maximum likelihood to estimate variance components for

animal models with several random effects using a derivative-free algorithm. *Genetics Selection Evolution, 21*(3), 1-24.

Meyer, K. (1990). *Present status of knowledge about statistical procedures and algorithms to estimate variance and covariance components.* Paper presented at the Proc 4th World Congress on Genetics Applied to Livestock Production, Edinburgh.

Meyer, K. (1998). Estimating covariance functions for longitudinal data using a random regression model. *Genetics Selection Evolution, 30*(3), 1-20.

Meyer, K. (2003). *Random regression models for analyses of longitudinal data in animal breeding.* Paper presented at the Proceedings 54th Session of the International Statistical Institute.

Meyer, K. (2005). Random regression analyses using B-splines to model growth of Australian Angus cattle. *Genetics Selection Evolution, 37*(6), 1-28.

Meyer, K. (2006). *WOMBAT–A program for mixed model analyses by restricted maximum likelihood. User notes.'(Animal Genetic and Breeding Unit.* Paper presented at the University of New England: Armidale) Meyer K, Graser HU.

Meyer, K. (2007). WOMBAT—A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *Journal of Zhejiang University Science B, 8*(11), 815-821.

Meyer, K., & Hill, W. G. (1997). Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated'records by restricted maximum likelihood. *Livestock Production Science, 47*(3), 185-200.

Meyer, K., & Houle, D. (2013). *Sampling based approximation of confidence intervals for functions of genetic covariance matrices.* Paper presented at the Proc Assoc Advmt Anim Breed Genet.

Meyer, K., & Kirkpatrick, M. (2005). Up hill, down dale: quantitative genetics of curvaceous traits. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1459), 1443-1455.

Miller, A. (2002). *Subset selection in regression*: CRC Press.

Miller, R. G. (1974). The jackknife-a review. *Biometrika, 61*(1), 1-15.

Misztal, I. (2006). Properties of random regression models using linear splines. *Journal*

*of Animal Breeding and Genetics, 123*(2), 74-80.

Misztal, I. (2008). Reliable computing in estimation of variance components. *Journal of Animal Breeding and Genetics, 125*(6), 363-370.

Morris, C. N. (1983). Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association, 78*(381), 47-55.

Morris, J. S. (2002). The BLUPs are not “best” when it comes to bootstrapping. *Statistics & Probability Letters, 56*(4), 425-430.

Morrison, D. F., Marshall, L. C., & Sahlin, H. L. (1976). Multivariate statistical methods.

Mrode, R. A. (2014). *Linear models for the prediction of animal breeding values*: Cabi.

Navarro, A., Zamorano, M. J., Hildebrandt, S., Ginés, R., Aguilera, C., & Afonso, J. M. (2009). Estimates of heritabilities and genetic correlations for body composition traits and G× E interactions, in gilthead seabream (Sparus auratus L.). *Aquaculture, 295*(3-4), 183-187.

Navarro, A., Zamorano, M. J., Hildebrandt, S., Ginés, R., Aguilera, C., & Afonso, J. M. (2009). Estimates of heritabilities and genetic correlations for growth and carcass traits in gilthead seabream (Sparus auratus L.), under industrial conditions. *Aquaculture, 289*(3-4), 225-230.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal, 7*(4), 308-313.

Ng, H. H., & Kottelat, M. (2008). The identity of Clarias batrachus (Linnaeus, 1758), with the designation of a neotype (Teleostei: Clariidae). *Zoological Journal of the Linnean Society, 153*(4), 725-732.

Nguyen, N. H., Khaw, H. L., Ponzoni, R. W., Hamzah, A., & Kamaruzzaman, N. (2007). Can sexual dimorphism and body shape be altered in Nile tilapia (Oreochromis niloticus) by genetic means? *Aquaculture, 272*, S38-S46.

Nguyen, N. H., Ponzoni, R. W., Abu-Bakar, K. R., Hamzah, A., Khaw, H. L., & Yee, H. Y. (2010). Correlated response in fillet weight and yield to selection for increased harvest weight in genetically improved farmed tilapia (GIFT strain), Oreochromis niloticus. *Aquaculture, 305*(1-4), 1-5.

Nielsen, H. M., Ødegård, J., Olesen, I., Gjerde, B., Ardo, L., Jeney, G., & Jeney, Z. (2010). Genetic analysis of common carp (Cyprinus carpio) strains: I: Genetic

parameters and heterosis for growth traits and survival. *Aquaculture, 304*(1-4), 14-21.

Ocaña, J., El Halimi, R., de Villa, M. C. R., & Sánchez, J. A. (2005). *Bootstrapping repeated measures data in a nonlinear mixed-models context*: Universitat de Barcelona. Institut de Matemàtica [IMUB].

Oliveira, S. N. D., Oliveira, C. A. L. D., Alexandre Filho, L., Resende, E. K. D., Barrero, N. M. L., Kunita, N. M., ... & Ribeiro, R. P. (2014). Genetic parameters and morphometric characteristics of two generations from the GIFT strain of the Nile Tilapia. Semina Ci. agr., 3457-3468.

Oliveira, H., Brito, L., Lourenco, D., Silva, F., Jamrozik, J., Schaeffer, L., & Schenkel, F. (2019). Invited review: Advances and applications of random regression models: From quantitative genetics to genomics. *Journal of Dairy Science, 102*(9), 7664-7683.

Pante, M. J. R., Gjerde, B., McMillan, I., & Misztal, I. (2002). Estimation of additive and dominance genetic variances for body weight at harvest in rainbow trout, Oncorhynchus mykiss. *Aquaculture, 204*(3-4), 383-392.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika, 58*(3), 545-554.

Patterson, H. D., & Thompson, R. (1975). *Maximum likelihood estimation of components of variance.* Paper presented at the Proceeding Eight International Biometric Conference, 1975.

Peixoto, J. L., & Harville, D. A. (1986). Comparisons of alternative predictors under the balanced one-way random model. *Journal of the American Statistical Association, 81*(394), 431-436.

Piepho, H. P., & Möhring, J. (2006). Selection in cultivar trials—Is it ignorable? *Crop Science, 46*(1), 192-201.

Piepho, H. P., Möhring, J., Melchinger, A., & Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica, 161*(1-2), 209-228.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R news, 6*(1), 7-11.

Plummer, M., Best, N., Cowles, K., Vines, K., Sarkar, D., Bates, D., . . . Magnusson, A. (2018). Output analysis and diagnostics for MCMC, package'coda'. In.

Ponzoni, R. W., Hamzah, A., Tan, S., & Kamaruzzaman, N. (2005). Genetic parameters and response to selection for live weight in the GIFT strain of Nile tilapia (Oreochromis niloticus). Aquaculture, 247(1-4), 203-210.

Ponzoni, R. W., Nguyen, N. H., Khaw, H. L., Hamzah, A., Bakar, K. R. A., & Yee, H. Y. (2011). Genetic improvement of Nile tilapia (Oreochromis niloticus) with special reference to the work conducted by the WorldFish Center with the GIFT strain. *Reviews in Aquaculture, 3*(1), 27-41.

Powell, M. J. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The computer journal, 7*(2), 155-162.

Premachandra, H., Nguyen, N. H., Miller, A., D'Antignana, T., & Knibb, W. (2017). Genetic parameter estimates for growth and non-growth traits and comparison of growth performance in sea cages vs land tanks for yellowtail kingfish Seriola lalandi. *Aquaculture, 479*, 169-175.

Psutka, J. V., & Psutka, J. (2019). Sample size for maximum-likelihood estimates of Gaussian model depending on dimensionality of pattern space. *Pattern Recognition, 91*, 25-33.

Quenouille, M. H. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society. Series B (Methodological), 11*(1), 68-84.

Quinton, C. D., Kause, A., Koskela, J., & Ritola, O. (2007). Breeding salmonids for feed efficiency in current fishmeal and future plant-based diet environments. *Genetics Selection Evolution, 39*(4), 431-446.

Rameez, R., Jahageerdar, S., Chanu, T. I., Jayaraman, J., & Bangera, R. (2020). Genetic variation among full-sib families and the effect of non-genetic factors on growth traits at harvest in Clarias magur (Hamilton, 1822). *Aquaculture Reports, 18*, 100411.

Reagan, R. E., Pardue, G. B., & Eisen, E. J. (1976). Predicting selection response for growth of channel catfish. *The Journal of heredity, 67*(1), 49-53.

Reis Neto, R. V., Oliveira, C. A. L. d., Ribeiro, R. P., Freitas, R. T. F. d., Allaman, I. B., & Oliveira, S. N. d. (2014). Genetic parameters and trends of morphometric traits of GIFT tilapia under selection for weight gain. *Scientia Agricola, 71*(4), 259-265.

Rezk, M. A., Ponzoni, R. W., Khaw, H. L., Kamel, E., Dawood, T., & John, G. (2009). Selective breeding for increased body weight in a synthetic breed of Egyptian Nile tilapia, Oreochromis niloticus: response to selection and genetic parameters.

*Aquaculture, 293*(3-4), 187-194.

Robertson, A. (1977). The effect of selection on the estimation of genetic parameters. *Zeitschrift für Tierzüchtung und Züchtungsbiologie, 94*(1-4), 131-135.

Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical science, 6*(1), 15-32.

Roff, D. A. (2008). Comparing sire and dam estimates of heritability: jackknife and likelihood approaches. *Heredity, 100*(1), 32-38.

Roff, D. A., & Preziosi, R. (1994). The estimation of the genetic correlation: the use of the jackknife. *Heredity, 73*(5), 544-548.

Rutten, M. J., Bovenhuis, H., & Komen, H. (2005). Genetic parameters for fillet traits and body measurements in Nile tilapia (Oreochromis niloticus L.). *Aquaculture, 246*(1-4), 125-132.

Rutten, M. J., Komen, H., & Bovenhuis, H. (2005). Longitudinal genetic analysis of Nile tilapia (Oreochromis niloticus L.) body weight using a random regression model. *Aquaculture, 246*(1-4), 101-113.

Sae-Lim, P., Komen, H., Kause, A., Martin, K. E., Crooijmans, R., van Arendonk, J. A., & Parsons, J. E. (2013). Enhancing selective breeding for growth, slaughter traits and overall survival in rainbow trout (Oncorhynchus mykiss). *Aquaculture, 372*, 89-96.

Sahoo, S. K., Ferosekhan, S., Giri, S., & Swain, S. (2016). Recent trends in breeding and seed production of Magur in India. *World Aquaculture Magazine, 47*(2), 59-62.

Sahoo, S. K., Ferosekhan, S., Giri, S. S., Radhakrishnan, K., Panda, D., SriHari, M., & Pillai, B. R. (2019). Length–weight relationship and growth performance of different life stages of hatchery-produced magur, Clarias magur (Hamilton, 1822). *Aquaculture Research, 50*(5), 1431-1437.

Saillant, E., Dupont-Nivet, M., Haffray, P., & Chatain, B. (2006). Estimates of heritability and genotype–environment interactions for body weight in sea bass (Dicentrarchus labrax L.) raised under communal rearing conditions. *Aquaculture, 254*(1-4), 139-147.

Schaeffer, L. (1984). Sire and cow evaluation under multiple trait models. *Journal of Dairy Science, 67*(7), 1567-1580.

Schaeffer, L. (2016). *Random regression models*. In. Retrieved from http://animalbiosciences. uoguelph. ca/~ lrs/BOOKS/rrmbook. pdf

Schaeffer, L. (2018). Necessary changes to improve animal models. *Journal of Animal Breeding and Genetics, 135*(2), 124-131.

Schaeffer, L. (2019). *Animal models*: Volumes Direct.

Schaeffer, L., & Dekkers, J. (1994). *Random regressions in animal models for test-day production in dairy cattle.* Paper presented at the World Congress of Genetics Applied Livestock Production, 1994.

Schenkel, F. S., Schaeffer, L. R., & Boettcher, P. J. (2002). Comparison between estimation of breeding values and fixed effects using Bayesian and empirical BLUP estimation under selection on parents and missing pedigree information. *Genetics Selection Evolution, 34*(1), 41-59.

Schweiger, R., Kaufman, S., Laaksonen, R., Kleber, M. E., März, W., Eskin, E., . . . Halperin, E. (2016). Fast and accurate construction of confidence intervals for heritability. *The American Journal of Human Genetics, 98*(6), 1181-1192.

Searle, S. R. (1971). A biometrics invited paper. Topics in variance component estimation. *Biometrics, 27*(1), 1-76.

Searle, S. R. (1995). An overview of variance component estimation. *Metrika, 42*(1), 215-230.

Searle, S. R. (1997). The matrix handling of BLUE and BLUP in the mixed linear model. *Linear algebra and its applications, 264*, 291-311.

Searle, S. R., Casella, G., & McCulloch, C. (1992). Variance components john wiley and sons. *Inc. New York*.

Searle, S. R., Casella, G., & McCulloch, C. E. (2009). *Variance components* (Vol. 391): John Wiley & Sons.

Searle, S. R., & Gruber, M. H. (1971). *Linear models* (Vol. 10): Wiley Online Library.

Seber, G. A., & Lee, A. J. (2012). *Linear regression analysis* (Vol. 329): John Wiley & Sons.

Serfling, R. J. (2009). *Approximation theorems of mathematical statistics* (Vol. 162): John Wiley & Sons.

Smith, A., Cullis, B. R., & Thompson, R. (2005). The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *The Journal of Agricultural Science, 143*(6), 449-462.

Smith, A. F., & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological), 55*(1), 3-23.

Smith, S., & Graser, H.-U. (1986). Estimating variance components in a class of mixed models by restricted maximum likelihood. *Journal of Dairy Science, 69*(4), 1156-1165.

Snijders, T. A., & Bosker, R. (1999). Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling. In: London: Sage.

Sorensen, D., & Kennedy, B. (1984). Estimation of genetic variances from unselected and selected populations. *Journal of Animal Science, 59*(5), 1213-1223.

Sorensen, D., Gianola, D., & Gianola, D. (2002). Likelihood, Bayesian and MCMC methods in quantitative genetics.

Srimai, W., Koonawootrittriron, S., Chaivichoo, P., Manee-aphai, W., Phu-onnim, A., Koolboon, U., & Na-Nakorn, U. (2020). Selection response and genetic parameters for growth in North African catfish, Clarias gariepinus (Burchell, 1822). *Aquaculture, 518*, 734843.

Srimai, W., Koonawootrittriron, S., Manee-Aphai, W., Chatchaiphan, S., Koolboon, U., & Na-Nakorn, U. (2019). Genetic parameters and genotype-environment interaction for growth traits of North African catfish, Clarias gariepinus (Burchell, 1822). *Aquaculture, 501*, 104-110.

Srisapoome, P., Chatchaiphan, S., Bunnoy, A., Koonawootrittriron, S., & Na-Nakorn, U. (2019). Heritability of immunity traits and disease resistance of bighead catfish, Clarias macrocephalus Günther, 1864. *Fish & shellfish immunology, 92*, 209-215.

Stroup, W., & Mulitze, D. (1991). Nearest neighbor adjusted best linear unbiased prediction. *The American Statistician, 45*(3), 194-200.

Stroup, W. W., Milliken, G. A., Claassen, E. A., & Wolfinger, R. D. (2018). *SAS for mixed models: introduction and basic applications*: SAS Institute.

Swallow, W. H., & Searle, S. R. (1978). Minimum variance quadratic unbiased estimation (MIVQUE) of variance components. *Technometrics, 20*(3), 265-272.

Tacon, A. G. (2018). Global trends in aquaculture and compound aquafeed production. *The Magazine of the World Aquaculture Society, 49*, 33-46.

Thai, H. T., Mentré, F., Holford, N. H., Veyrat-Follet, C., & Comets, E. (2013). A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. *Pharmaceutical statistics, 12*(3), 129-140.

Thompson, R. (1973). The estimation of variance and covariance components with an application when records are subject to culling. *Biometrics*, 527-550.

Thompson, R. (1979). Sire evaluation. *Biometrics*, 339-353.

Thompson, R., Brotherstone, S., & White, I. M. (2005). Estimation of quantitative genetic parameters. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1459), 1469-1477.

Thompson, R., & Mantysaari, E. (1999). Prospects for statistical methods in dairy cattle breeding. *Proceedings of the Genetic Improvement of Functional Traits in cattle longevity, Interbull. Jouy-en-Jossas, France, 1999, 20*, 71-78.

Thompson, R., & Meyer, K. (1986). A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livestock Production Science, 15*(4), 299-313.

Turra, E. M., de Oliveira, D. A. A., Valente, B. D., de Alencar Teixeira, E., de Assis Prado, S., de Melo, D. C., e Silva, M. d. A. (2012). Estimation of genetic parameters for body weights of Nile tilapia Oreochromis niloticus using random regression models. *Aquaculture, 354*, 31-37.

Van der Werf, J. (2002). *Optimising selection for traits along trajectories.* Paper presented at the Seventh World Congress on Genetics Applied to Livestock Production. Communication.

Van der Werf, J., & De Boer, I. (1990). Estimation of additive genetic variance when base populations are selected. *Journal of Animal Science, 68*(10), 3124-3132.

Van der Werf, J., Goddard, M., & Meyer, K. (1998). The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. *Journal of Dairy Science, 81*(12), 3300-3308.

Van der Werf, J., & Schaeffer, L. (1997). Random regression in animal breeding. *Ontario: University of Guelph*, p70.

Van Tassell, C., & Van Vleck, L. D. (1996). Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co) variance component inference. *Journal of Animal Science, 74*(11), 2586-2597.

Vandeputte, M., Kocour, M., Mauger, S., Rodina, M., Launay, A., Gela, D., Linhart, O. (2008). Genetic variation for growth at one and two summers of age in the common carp (Cyprinus carpio L.): Heritability estimates and response to selection. *Aquaculture, 277*(1-2), 7-13.

Vishwanath, W. 2010. Clarias magur. The IUCN Red List of Threatened Species 2010: e.T168255A6470089.https://dx.doi.org/10.2305/IUCN.UK.2010-4.RLTS.T168255A6470089.en. Downloaded on 14 January 2021.

Vu, N. T., Van Sang, N., Phuc, T. H., Vuong, N. T., & Nguyen, N. H. (2019). Genetic evaluation of a 15-year selection program for high growth in striped catfish Pangasianodon hypophthalmus. *Aquaculture, 509*, 221-226.

Wade, K., Quaas, R., & Van Vleck, L. D. (1993). Estimation of the parameters involved in a first-order autoregressive process for contemporary groups. *Journal of Dairy Science, 76*(10), 3033-3040.

Waldmann, P., & Ericsson, T. (2006). Comparison of REML and Gibbs sampling estimates of multi-trait genetic parameters in Scots pine. *Theoretical and Applied Genetics, 112*(8), 1441-1451.

Walsh, B. (2001). Quantitative genetics in the age of genomics. *Theoretical Population Biology, 59*(3), 175-184.

Walsh, B., & Lynch, M. (2018). *Evolution and selection of quantitative traits*: Oxford University Press.

Wegman, E. J., & Wright, I. W. (1983). Splines in statistics. *Journal of the American Statistical Association, 78*(382), 351-365.

White, I., Thompson, R., & Brotherstone, S. (1999). Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Dairy Science, 82*(3), 632-638.

Wiggans, G., Goddard, M., & Meyer, K. (1996). Test day model with 30 traits and a genetic (co) variance matrix of reduced rank. *J. Dairy Sci.*

Winkelman, A., & Peterson, R. (1994). Genetic parameters (heritabilities, dominance ratios and genetic correlations) for body weight and length of chinook salmon after

9 and 22 months of saltwater rearing. *Aquaculture, 125*(1-2), 31-36.

Yang, R.-C. (2010). Towards understanding and use of mixed-model analysis of agricultural experiments. *Canadian Journal of Plant Science, 90*(5), 605-627.

Zhao, J., Zhao, Y., Song, Z., Liu, H., Liu, Y., & Yang, R. (2018). Genetic analysis of the main growth traits using random regression models in Japanese flounder (Paralichthys olivaceus). *Aquaculture Research, 49*(4), 1504-1511.

**Limitations of present study**

- The genetic parameters estimated in the present study has the limitation due to the mating design employed.
- Non-standardization of management practices especially with regard to survival at early stages biased the estimates further.
- Lack of standardized management practices especially the reproduction/breeding aspects and larval rearing resulted in the production of limited number of families and number of offspring per family leading to reduced sample size.
- Failure of the stocks to reproduce resulted in loss of genetic material in the formation of base population with large genetic base.

**Future directions**

- In future it will be desirable to use nested and/or factorial mating designs to estimate unbiased genetic parameters.
- Also, it will be beneficial to have more number of families and family replications to partition the non-additive genetic factors.
- The application of molecular pedigree assigning may be used to reduce the common environmental effect and generation interval.

# R_Codes

## 1. Non-Parametric Bootstraping

```
fixed<-master$fi
animal<-master$ai
error<-master$ei
dam<-master$Damid
common<-master[,c(-9:-12)]
anidam<-as.data.frame(cbind(animal,dam))
errdam<-as.data.frame(cbind(error,dam))
size<-(as.data.frame(table(dam)))$Freq
zz<-NULL
for(i in 1:10000)
{
    newy<-fixed+(getdata(anidam,(strata(anidam, "dam", size = size, method =
"srswr")))$animal)+
    (getdata(errdam,(strata(errdam, "dam", size = size, method = "srswr")))$
error)
    boot_dat<-cbind(common,newy)
    write.table(boot_dat,"boot_dat.d",row.names=FALSE,col.names = FALSE)
     system("wombat.exe wombat.par")
    z<-get_lines("SumEstimates.out", c(44, 55))
    zz<-rbind(zz,z)
}
```

## 2. Parametric Bootstrap

```
write.table(boot,"boot.d",row.names=FALSE,col.names = FALSE)
zz<-NULL
for(i in 1:10000)
{
  system("wombat.exe wombat.par")
  system("wombat.exe wombat1.par")
  z<-get_lines("SumEstimates.out", c(44, 55))
  zz<-rbind(zz,z)
}
write.table(zz,"zzout.txt")
```

## 3. Bayesian MCMC

```
write.table(mped,"mped.txt")
pedigree <- read.table("harvped.txt", header = T)
write.table(harv,"harv.txt")
data <- read.table("harv.txt", header = T)
prior <- list(R = list(V=1, nu=0.002), G = list(G1 = list(V=1, nu=0.002)))
modelharv <- MCMCglmm(BW5 ~ 1+BW1+stock+Batch+Pond+Sex, random = ~animal,fam
ily = "gaussian",
```

```
prior = prior, pedigree = pedigree, data = data, nitt = 1000000,burnin = 500
00, thin = 100)
summary(model)
herit <- model$VCV[, "Sireid"]+model$VCV[, "Sireid"]/(model$VCV[, "Sireid"]
+ model$VCV[, "units"])
herit <- model$VCV[, "animal"]/(model$VCV[, "animal"] + model$VCV[, "units"]
)
mean(herit)
effectiveSize(herit)
```

### 4. Jackknife

```
zz<-NULL
n=1
for (i in 1:1413)
{
  harv1<-harv[-n,]
  n=n+1
  write.table(harv1,"harv.d",row.names = F,col.names = F,quote = F)
  system("wombat.exe wombat.par")
  z<-get_lines("SumEstimates.out", c(44, 55))
  zz<-rbind(zz,z)
}
write.table(zz,"zzout.txt",row.names=FALSE,col.names = FALSE,quote = F)
```

###R-Plots

```
rr1<-rr[rr$Animal %in% rr$Animal[duplicated(rr$Animal)],]

ggplot(rr1[rr1$Batch==2,], aes(x = age, y = bw))+
  geom_line(aes(color = Animal, group = Animal))+ theme_bw()+
  theme(legend.position = "none",strip.background = element_blank(),
        strip.text =  element_blank()) +
  geom_smooth(color='yellow')+
  #stat_summary(geom="line",fun = mean,size=1,color='yellow') +
  scale_colour_gradient(low="grey", high="steelblue")+
facet_wrap(~Pond)

ggplot(rr1[with(rr1, Batch == 2 & Sex == 1), ], aes(x = age, y = bw))+
  geom_line(aes(color = Animal, group = Animal))+ theme_bw()+
  theme(legend.position = "none",strip.background = element_blank(),
        strip.text =  element_blank()) +
  geom_smooth(color='yellow')+
  scale_colour_gradient(low="grey", high="steelblue")

ggplot(fs1, aes(Family, ADG, color = Model)) +
  geom_segment(aes(x=Family, xend=Family, y=0, yend=ADG),color="blue") +
  geom_point()+theme_bw()
```

```
ggplot(fs, aes(y=Family)) +
  geom_dumbbell(aes(x=ADG_Uni, xend=ADG_Multi),size=0.8, color="#e3e2e1",
                colour_x = "red", colour_xend = "blue") + xlab("Accuracy")+
  ylab("Family ID")+
  geom_text(x=.125, y=75, label="Multivariate_ADG",color="blue",size=3)+
  geom_text(x=.075, y=75, label="Univariate_ADG", color="red",size=3)+
  theme_bw() + ggtitle("Accuracies of Univariate and Multivariate Family BVs
")+
  theme(plot.title = element_text(hjust = 0.5,size=10, face="bold"))+
  coord_flip()+xlim(c(0:1))

ggplot(fs) +
  #geom_line(aes(x = ADG, y = Model, group = Family, color=Model),size=1) +
  geom_point(aes(x = Family, y = K, color = Model), size = 1.5) +
  theme_bw()+ylim(c(0:1))

ggplot(vrat2, aes(x=Age, y=Heritability, group=Model)) +
  geom_line(aes(linetype=Model,color=Model))+
  geom_point(aes(shape=Model,color=Model))+
  #geom_smooth(aes(color=Model))+
  theme_bw()+ylim(c(0:1))

ggplot(hetf, aes(as.factor(Age1), as.factor(Age2), fill= PEnv)) +
  geom_tile()+
  #scale_fill_gradient(low="red", high="green")+
  scale_fill_distiller(palette = "Set1")

  ggplot(hom, aes(Age1,Age2,  z=PE)) +
    geom_contour_filled(binwidth = 0.10005)+
    #scale_fill_brewer(palette = "mycolors")+
    scale_fill_manual(values = colorRampPalette(brewer.pal(9, "RdPu"))(13))
+
    theme_bw() + theme(legend.key.height = unit(4.5,"mm") )

# Static chart
plot3d(hom$Age1, hom$Age2, hom$Genetic, type = "p", radius = 1)
play3d( spin3d( axis = c(0, 0, 1), rpm = 100), duration = 10 )


# We can indicate the axis and the rotation velocity
play3d( spin3d( axis = c(0, 0, 1), rpm = 20), duration = 10 )

ggplot(efhom, aes(x=Age, y=Eigenfunction, group=Variate)) +
  geom_line(aes(linetype=Variate,color=Variate))+
  geom_point(aes(shape=Variate,color=Variate))+
  geom_hline(yintercept=0,linetype="dashed", color = "red")+
  #geom_smooth(aes(color=Model))+
  theme_bw()#+ylim(c(-2,2))
```

# SAS_Codes

## 1. SAS Proc Mixed for Animal model

```
***ANIMAL MODEL FOR HERITABILITY AND SE;
proc mixed data=harv covtest asycov;
class animalid stock batch pond sex;
model bw5 = bw1 stock batch pond(batch) sex;
random animalid/type=lin(1) ldata=L2data ;
ods output covparms=_varcomp asycov=_cov;
run;
PROC IML;
use _varcomp;
read all var {Estimate} into _varcomp;
close _varcomp;
use _cov;
read all var {CovP1 CovP2 } into _cov;
close _cov;
A={1 0 }*_varcomp ; /* Additive genetic variance */
P={1 1 }*_varcomp ; /* Phenotypic variance */
E={0 1 }*_varcomp;  /*Error variance */
h2_i=A/P;           /* Heritability */
c_n={1, 0};         /* Coefficients of numerator (Additive var) */
c_d = {1,1};        /* Coefficients of denominator (Pheno. var) */
var_A =c_n` * _cov * c_n ; /* Variance of Additive variance */
var_P = c_d` * _cov * c_d ; /* Variance of Phenotypic variance */
cov_A_P = c_n` * _cov * c_d ; /* Covariance between Additive and Phenotypic
*/
var_h2_i_dick = var_A / (P**2); /* Variance of heritability: Dickerson */
se_h2_i_dick = SQRT(var_h2_i_dick) ; /* Std Error of heritability: Dickerson
*/
var_h2_i_delta=(h2_i**2)*((var_A/A**2)+(var_P/P**2)-
(2*cov_A_P/(A*P))); /* Variance of heritability: Delta method */
SE_h2_i_delta = SQRT(var_h2_i_delta) ; /* Std Error of heritability: Delta m
ethod */
SE_A=sqrt(var_A) ;
PRINT
A P E
h2_i [format=6.2]
SE_h2_i_delta [format=6.3]
SE_h2_i_dick [format=6.3];
QUIT;
```

## 2. SAS Proc Mixed for Fullsib model

```
proc mixed data=magur method=type2 covtest asycov;
class sireid batch pondno gender;
model bw5 = taggingage batch pondno(batch) gender;
```

```
random sireid(batch);
ods output covparms=_varcomp asycov=_cov;
run;
PROC IML;
use _varcomp;
read all var {Estimate} into _varcomp;
close _varcomp;
use _cov;
read all var {CovP1 CovP2 } into _cov;
close _cov;
A={2 0 }*_varcomp ; /* Additive genetic variance */
P={1 1 }*_varcomp ; /* Phenotypic variance */
E={0 1 }*_varcomp;  /*Error variance */
h2_i=A/P;           /* Heritability */
c_n={2, 0};         /* Coefficients of numerator (Additive var) */
c_d = {1,1};        /* Coefficients of denominator (Pheno. var) */
var_A =c_n` * _cov * c_n ; /* Variance of Additive variance */
var_P = c_d` * _cov * c_d ; /* Variance of Phenotypic variance */
cov_A_P = c_n` * _cov * c_d ; /* Covariance between Additive and Phenotypic
*/
var_h2_i_dick = var_A / (P**2); /* Variance of heritability: Dickerson */
se_h2_i_dick = SQRT(var_h2_i_dick) ; /* Std Error of heritability: Dickerson
*/
var_h2_i_delta=(h2_i**2)*((var_A/A**2)+(var_P/P**2)-
(2*cov_A_P/(A*P))); /* Variance of heritability: Delta method */
SE_h2_i_delta = SQRT(var_h2_i_delta) ; /* Std Error of heritability: Delta m
ethod */
SE_A=sqrt(var_A) ;
PRINT
A P E
h2_i [format=6.2]
SE_h2_i_delta [format=6.3]
SE_h2_i_dick [format=6.3];
QUIT;
```

### 3. Coverage Probability

```
proc surveyselect data=all out=sim seed=224699 method=srs sampsize=50 reps=2
00 ;
run;
proc means data=sim noprint;
   by replicate;
   var additiveNPB residualNPB heritabilitynpb additivepb residualpb heritab
ilitypb additivebys residualbys
   heritabilitybys additiveasy residualasy heritabilityasy;
   output out=OutStats
   mean=mean= Manpb Mrnpb Mhnpb Mapb Mrpb Mhpb Mabys Mrbys MhBys Maasy Mrasy
Mhasy
   lclm=anpbL rnpbL hnpbL apbL rpbL hpbL abysL rbysL hbysL aasyL rasyL hasyL
   uclm=anpbU rnpbU hnpbU apbU rpbU hpbU abysU rbysU hbysU aasyU rasyU hasyU
;
```

```
run;
data OutStats;  set OutStats;
   *label ParamInCI = "Parameter in CI";
   ParamInANPB = (anpbL<240   & anpbU>240);           /* indicator variable
*/
   ParamInRNPB = (rnpbL<230   & rnpbU>230);
   ParamInHNPB = (HnpbL<0.51  & hnpbU>0.51);
   ParamInAPB  = (apbL<255.6  & apbU>255.6);
   ParamInRPB  = (rpbL<325.2  & rpbU>325.2);
   ParamInHPB  = (hpbL<0.44   & hpbU>0.44);
   ParamInABYS = (abysL<279.4 & abysU>279.4);
   ParamInRBYS = (rbysL<395.4 & rbysU>395.4);
   ParamInHBYS = (hbysL<0.41  & hbysU>0.41);
   ParamInAASY = (aasyL<257.6 & aasyU>257.6);
   ParamInRASY = (rasyL<325.7 & rasyU>325.7);
   ParamInHASY = (hasyL<0.44  & hasyU>0.44);
run;
proc freq data=OutStats;
   tables ParamInANPB / nocum binomial(level='1' p=0.95);
   tables ParamInRNPB / nocum binomial(level='1' p=0.95);
   tables ParamInHNPB / nocum binomial(level='1' p=0.95);
   tables ParamInAPB / nocum binomial(level='1' p=0.95);
   tables ParamInRPB / nocum binomial(level='1' p=0.95);
   tables ParamInHPB / nocum binomial(level='1' p=0.95);
   tables ParamInABYS / nocum binomial(level='1' p=0.95);
   tables ParamInRBYS / nocum binomial(level='1' p=0.95);
   tables ParamInHBYS / nocum binomial(level='1' p=0.95);
   tables ParamInAASY / nocum binomial(level='1' p=0.95);
   tables ParamInRASY / nocum binomial(level='1' p=0.95);
   tables ParamInHASY / nocum binomial(level='1' p=0.95);

run;
proc univariate data=sim noprint;
   by replicate;
   var additiveNPB residualNPB heritabilitynpb additivepb residualpb heritab
ilitypb additivebys residualbys
   heritabilitybys additiveasy residualasy heritabilityasy;
   output out=OutStats mean= Manpb Mrnpb Mhnpb Mapb Mrpb Mhpb Mabys Mrbys Mh
Bys Maasy Mrasy Mhasy
            pctlpts  = 2.5 97.5
                    pctlpre  = anpb rnpb hnpb apb rpb hpb abys rbys hbys aas
y rasy hasy
                    pctlname = L U;
run;
```

## 4. Kernal Density Plots

```
ods graphics on;
proc kde data=magur;
   bivar totallength bodyweight / plots=all;
run;
```

```
ods graphics off;
ods graphics on;

proc kde data=all;
univar Additive_AM_Scheme1  Additive_AM_Scheme2 Additive_AM_Scheme3 Additive
_Asy
Additive_Bayes  Additive_FM_Scheme1 Additive_FM_Scheme2
    Additive_FM_Scheme3 Additive_PB/ plots=densityoverlay;
univar Residual_AM_Scheme1  Residual_AM_Scheme2 Residual_AM_Scheme3 Residual
_Asy
Residual_Bayes  Residual_FM_Scheme1 Residual_FM_Scheme2
    Residual_FM_Scheme3 Residual_PB/plots=densityoverlay;
univar h2_AM_Scheme1    h2_AM_Scheme2   h2_AM_Scheme3   h2_Asy  h2_Bayes
h2_FM_Scheme1
h2_FM_Scheme2   h2_FM_Scheme3   h2_PB/plots=densityoverlay;
run;

proc kde data=density;
univar var_family1 var_residual1 var_additive1 var_environmental1 heritabili
ty1
 / plots=(density histogram histdensity);
 univar var_family2 var_residual2 var_additive2 var_environmental2 heritabil
ity2
 / plots=(density histogram histdensity);
univar var_family1 var_residual1  / plots=densityoverlay;
univar var_additive1 var_environmental1  / plots=densityoverlay;
univar var_family1 var_additive1  / plots=densityoverlay;
univar var_residual1 var_environmental1  / plots=densityoverlay;
univar var_family1 var_residual1 var_additive1 var_environmental1 / plots=de
nsityoverlay;
univar var_family2 var_residual2  / plots=densityoverlay;
univar var_additive2 var_environmental2  / plots=densityoverlay;
univar var_family2 var_residual2 var_additive2 var_environmental2 / plots=de
nsityoverlay;
univar var_family2 var_additive2  / plots=densityoverlay;
univar var_residual2 var_environmental2  / plots=densityoverlay;
univar var_family1 var_family2  / plots=densityoverlay;
univar var_residual1 var_residual2 / plots=densityoverlay;
univar var_additive1 var_additive2  / plots=densityoverlay;
univar var_environmental1 var_environmental2  / plots=densityoverlay;
univar heritability1 heritability2  / plots=densityoverlay;
run;
ods graphics off;
proc iml;
use post;
read all var {Var_Dam Var_Residual Var_Additive Var_Environmental herit1 } i
nto x;
close post;
use post1;
read all var {Var_Dam Var_Residual Var_Additive Var_Environmental herit2 } i
```

```
nto y;
close post1;
herit=x||y;
CREATE density FROM herit [COLNAME={Var_Family1 Var_Residual1 Var_Additive1
Var_Environmental1 Heritability1
Var_Family2 Var_Residual2 Var_Additive2 Var_Environmental2 Heritability2}];
APPEND FROM herit;
quit;
```

# Wombat_Codes

## 1. Univariate animal model - Wombat

```
COM Simple univariate analysis

# specify the pedigree file to be used
PEDS ../harv.d

# specify the data file and its layout
DATA ../harv.d
  animal
  sire 76
  dam 78
  stock 2
  culturetype 2
  age
  batch 2
  pond 11
  sex 2
  culturedays
  nbw1
  tl1
  nbw5
  tl5
  lntl5
  bd5
  lnbd5
  hw5
  lnhw5
  k5
  adg
end

# choose the type of analysis
ANAL UNI

# specify the model to be fitted
MODEL
 FIX animal
```

```
 FIX pond
 FIX stock
 FIX batch
 FIX sex
 COV nbw1(1)
 TR nbw5
END

# give starting values for variance compoents

VAR residual 1
325.89
```

## 2. Multivariate animal model -Wombat

```
RUNOP  --pivot0.000000002 --batch
COM  Hexavariate analysis of bw5 tl5 bd5 hw5 k5 adg at harvest
ANAL MUV 6
PEDS ../harv.ped
DATA ../harvhexa.d GRP
  TRNOS 1 2 3 4 5 6
  traitno    6
  animal     0
  sire
  dam
  stock      2
  culturetype 2
  age
  batch      2
  pond       11
  sex        2
  culturedays
  bw1
  tl1
  NAMES bw5 tl5 bd5 hw5 k5 adg
END

MODEL
  FIX stock
  FIX batch
  FIX pond
  FIX sex
  COV bw1(1)
  RAN animal  nrm 1 2 3 4 5 6
  trait bw5  1
  tr    tl5  2
  tr    bd5  3
  tr    hw5  4
  tr    k5   5
  tr    adg  6
END MOD
```

```
VAR animal 6
       255.72000000
        14.76100000
        2.26070000
        1.95380000
        0.34741000
        0.58042000
        0.97873000
        0.11957000
        0.11549000
        0.00950800
        0.03773500
        0.02373200
        0.01860200
        0.00410420
        0.00444890
        0.02346400
        0.00284930
        0.00471060
        0.00175270
        0.00084744
        0.00232400

VAR residual 6
        324.860000000
          17.882000000
          2.172100000
          2.829800000
          0.134420000
          0.625870000
          2.065500000
          0.151280000
          0.203490000
          -0.131360000
          0.048687000
          0.082482000
          0.029909000
          -0.005442000
          0.006530400
          0.063863000
          -0.006315600
          0.007905100
          0.023949000
          0.000400840
          0.002685800

#SPECIAL
#PENALTY PACORR PHENV animal 8.0
```

```
#PENALTY PACORR PHENV residual 8.0
#END
```

### 3. Random regression homogeneous variance model - Wombat

```
RUNOP --logdiag --pxai150,50
COM  Multivariate analysis of Body weight at five different time points
ANAL RR
PEDS ../RRfull.ped
DATA ../RRlatest1.d
  number
  animal
  sire
  dam
  subject 9999
  family 78
  batch  2
  stock  3
  culturetype 2
  pond   11
  sex  2
  sampling 5
  age 99
  age1 99
  bw
END

MODEL
  FIX batch
  FIX stock
  FIX pond
  FIX sex
  COV batch*age(3,leg)
  RRC age
  RAN animal(3,leg) nrm
  RAN subject(3,leg)
  trait bw
END MOD

# give starting values for covariances: upper triangle!
VAR animal 3
704.147
136.842
-107.979
161.042
-45.1949
53.9441
VAR subject 3
180.1396
297.7924
139.2841
```

```
660.6786
311.2506
159.1818
# fit heterogeneous error variances with 6 classes
VAR residual 1 HOM 1
50

SPECIAL
# force correlations amongst all ages to be written out
  RRCORR-ALL
END
```

### 4. Random regression heterogeneous variance model - Wombat

```
RUNOP --logdiag --pxai150,50
COM  Multivariate analysis of Body weight at five different time points
ANAL RR
PEDS ../RRfull.ped
DATA ../RRlatest1.d
  number
  animal
  sire
  dam
  subject 9999
  family 78
  batch  2
  stock  3
  culturetype 2
  pond   11
  sex  2
  sampling 5
  age 99
  age1 99
  bw
END

MODEL
  FIX batch
  FIX stock
  FIX pond
  FIX sex
  COV batch*age(3,leg)
  RRC age
  RAN animal(3,leg) nrm
  RAN subject(3,leg)
  trait bw
END MOD

# give starting values for covariances: upper triangle!
VAR animal 3
704.147
```

```
136.842
-107.979
161.042
-45.1949
53.9441
VAR subject 3
180.1396
297.7924
139.2841
660.6786
311.2506
159.1818
#VAR family 3
#397.39
#190.27
#40.8799
#125.92
#13.788
#30.826
# fit heterogeneous error variances with 6 classes
VAR residual 1 HET 50
180 189  40.28585063
190 199  65.1380771
200 209  40.61546287
210 219  30.80349458
220 270  49.08299794
271 280  99.73152798
281 290  50.04482793
291 300  20.32508738
301 310  56.73308792
311 319  80.81742791
320 329  40.33949308
330 339  59.93994471
340 349  40.73645577
350 359  55.6305192
360 361  45.61728783
362 371  70.59952239
372 381  24.271162
382 391  48.5059673
392 401  34.85551929
402 410  46.87570382
411 420  26.39669659
421 430  48.28579826
431 440  29.00647081
441 450  55.10612747
451 452  41.8093227
453 462  85.42834986
463 472  23.10432499
473 482  36.00521171
483 492  32.21877353
```

```
493 501  35.95442684
502 511  70.55333395
512 521  77.23807804
522 531  74.46462911
532 541  64.87697089
542 543  96.42084096
544 553  90.51581537
554 563  37.06020012
564 573  78.47172918
574 583  57.7663089
584 592  48.80793061
593 602  35.87830715
603 612  98.95813817
613 622  82.85244495
623 632  86.68510288
633 683  45.33494404
684 693  37.1217596
694 703  44.21233824
704 713  58.44629906
714 723  83.57717739
724 725  27.6077983

SPECIAL
# force correlations amongst all ages to be written out
  RRCORR-ALL
END
```