

---

---

# ВЕТЕРИНАРИЯ И ЗООТЕХНИЯ

DOI: 10.31677/2072-6724-2022-65-4-77-96

УДК 575.167

## ПРОБЛЕМЫ И ВОПРОСЫ ПРИ ПРОГНОЗИРОВАНИИ ГЕНЕТИЧЕСКОЙ ПЛЕМЕННОЙ ЦЕННОСТИ СЕЛЬКОХОЗЯЙСТВЕННЫХ ЖИВОТНЫХ

<sup>1,2</sup>А.Е. Калашников, кандидат биологических наук<sup>1</sup>А.И. Голубков, доктор сельскохозяйственных наук, профессор<sup>1</sup>Н.Ф. Щегольков, кандидат сельскохозяйственных наук, доцент<sup>3</sup>Е.Р. Гостева, доктор сельскохозяйственных наук<sup>1</sup>*Всероссийский научно-исследовательский институт племенного дела Минсельхоза России, Москва, Россия*<sup>2</sup>*ФИЦ комплексного изучения Арктики им. акад. Н.П. Лаверова УРО РАН, Архангельск, Россия*<sup>3</sup>*Федеральный аграрный научный центр Юго-Востока, Саратов, Россия*

E-mail: ekagosteva@yandex.ru

**Ключевые слова:** племенная ценность, Хендерсон, прогнозирование факторов, линейные модели, смешанные модели, BLUP.

**Реферат.** Методы прогнозирования генетической ценности в скотоводстве получили широкое признание в странах с развитым животноводством, они используются и совершенствуются в непрерывной эволюции как генетиками-селекционерами, так и статистиками. Метод несмещенной линейной оценки (BLUP/BLUE) является очень гибким, так как в нем существует множество альтернатив, которые приспособлены к различным селекционным целям, видам животных, условиям производственной среды и способам вычислений. Сегодня актуален поиск новых, более быстрых и экономичных алгоритмов инвертирования доминантных и аддитивных родственных взаимоотношений между индивидуумами с помощью аддитивных матриц родства. Для этого необходимо выбрать (создать) подходящую модель отбора для решения проблемы выбраковки коров, неслучайного влияния быков-производителей, увязать это с фиксированными факторами жизненного цикла животного HYS (стадо – год – сезон, herd – year – season) и факторами предпочтительного производственного окружения. Целью данной работы является рассмотрение ряда вопросов, касающихся проблем разведения животных. Представлен обзор исторически значимого метода оценки генетической ценности крупного рогатого скота (и других видов по аналогии), основанного на многофакторных регрессионных моделях, и приведены простые примеры отбора при его помощи производителей. За последнее десятилетие опубликовано множество новых методических приемов в этой области, новых программ, баз данных, патентов, обзорных статей. В соответствии с требованиями национального экономического развития представленный материал может служить руководством к изучению новых, современных методов оценки ценности животных и формированию новых селекционных индексов.

## PROBLEMS AND ISSUES IN FORECASTING THE GENETIC BREEDING VALUE OF AGRICULTURAL ANIMALS

<sup>1,2</sup>A.E. Kalashnikov, Ph.D. in Biological Sciences<sup>1</sup>A.I. Golubkov, Doctor of Agricultural Sciences, Professor<sup>1</sup>N.F. Schegolkov, Ph.D. in Agricultural Sciences, Associate Professor<sup>3</sup>E.R. Gosteva, Doctor of Agricultural Sciences<sup>1</sup>*Federal State Budgetary Scientific Institution All-Russian Research Institute of Breeding of the Ministry of Agriculture of Russia, Moscow, Russia*<sup>2</sup>*N.P. Lavrov Federal Center for Integrated Arctic Research of the Ural Branch of the Russian Academy of Sciences, Arkhangelsk, Russia*<sup>3</sup>*Federal State Budgetary Scientific Organization «Federal Center of Agriculture Research of the South-East Region» Saratov, Russia*

E-mail: ekagosteva@yandex.ru

**Keywords:** breeding value, Henderson, forecasting of the factors, linear models, mixed models, BLUP.

**Abstract.** Methods for forecasting genetic value in cattle breeding are widely recognized in countries with developed cattle breeding. They are used and improved in continuous evolution by genetic breeders and statisti-

cians. The unbiased linear estimation method (BLUP/BLUE) is very flexible because it has many alternatives tailored to different breeding objectives, animal species, production conditions, and calculation methods. Today it is relevant to the search for new, faster, and more economical algorithms for inverting dominant and additive kinship relationships between individuals using additive kinship matrices. For a new search, it is necessary to select (create) a suitable selection model to solve the problem of cow culling, the non-random influence of sires. Further, it is essential to relate this to the fixed factors of the animal's life cycle HYS (herd-year-season, herd-year-season) and the preferred production environment factors. The purpose of this paper is to address several issues related to the problems of animal breeding. First, an overview of a historically powerful method for assessing the genetic value of cattle (and other species by analogy) based on multifactorial regression models is presented, and simple examples of selection using it are given. Over the past decade, many new methodological techniques, programs, databases, patents, and review articles have been published in this area. By the requirements of national economic development, the material presented can serve as a guide for the study of new, modern methods of assessing the value of animals and the formation of new breeding indices.

В свое время Ч. Хендерсон был вдохновлен статьей 1933 г. Дж. Лаша «Проблема бычьего индекса в свете современной генетики» [1], в которой были поставлены вопросы, изучение которых привело к идее метода BLUP (наилучший линейный несмещенный прогноз, best linear unbiased<sup>1</sup> prediction, который проводится методом BLUE, наилучшего линейного несмещенного вычисления, best linear unbiased equation) и методу оценки компонентов дисперсии, смещения оценки. Статья была проиллюстрирована рядом конкретных примеров и алгоритмов, и Ч. Хендерсон унаследовал в своих статьях этот прием изложения материала. Все его публикации содержат точные и последовательно сформулированные предположения и условия, для которых применимы модели и методология алгоритмов вычислений. Цель этой статьи состоит в том, чтобы изложить основные предположения Ч. Хендерсона в области генетической оценки племенной ценности, уделив особое внимание тем проблемам, которые еще предстоит решить в будущем. В статье также отмечены усовершенствования методов линейной оценки, которые в настоящее время предложены другими исследователями.

Что понимается под выбором переменных при оценке племенной ценности животных? Выбор переменных означает выбор фенотипических (инструментально измеренных) и генотипических исходных данных, проверку их достоверности, надежности, описание выборки при помощи методов описательной статистики, фильтрацию и подготовку в заданном формате для включения в линейную модель. Это именно работа с данными, а не

вариант выбора метода селекции, например, который сам по себе является переменной. Таким образом, выбор переменных – это частный случай выбора самой модели, написание первичных уравнений в скалярной или векторной форме, предположение о том, как будет выглядеть система уравнений ММЕ в матричной форме.

Классические методы выбора переменных включают прямую работу с фенотипическими данными, обратное исключение данных и итерационный метод. Прямой выбор данных может начинаться без использования выборки фенотипических переменных, которые были известны заранее. Краеугольным камнем здесь является изначальный выбор правильного подмножества фенотипических данных, когда нужно смоделировать взаимосвязь между интересующей переменной и подмножеством потенциальных объясняющих переменных или предикторов, но существует неопределенность в отношении того, какое подмножество использовать в целом.

Важность использования переменных исходных данных в селекции и оценке племенной ценности заключается в том, насколько данная математическая модель «использует» эту переменную для получения точных прогнозов племенной ценности. Чем больше модель оценивает значимость переменной для прогнозирования, тем важнее она в модели. Взвешивание переменных может применяться во многих линейных моделях, каждая из которых использует разные показатели оценки значимости исходных данных.

Важно понимать, что переменные являются основными единицами информации, из-

<sup>1</sup> *unbiased* (англ.) переводится как несмещенный, неискаженный, объективный, есть еще его противоположность, если регрессные коэффициенты в уравнениях не имеют своих весов, – смещенный, необъективный, искаженный прогноз *biased* (англ.) – BLBE, но он не применяется в животноводстве и промышленности, т. к. неэффективен относительно генеральной совокупности данных и требует обучения модели.

учаемой и интерпретируемой в оценке и корреляциях. В частности, размер номинальных выборок по достоверности в экспериментах имеет тенденцию к завышению, особенно если выборки сформированы под заранее выбранную модель. С другой стороны, выбор переменных не оказывает большого влияния на ошибку дисперсии, если его правильно сформировать относительно категорий данных и групп, стратегии кластеризации в селекции. Переменными для генетической и фенотипической оценки могут быть только числа или величины, которые можно измерить или подсчитать.

## ОБЪЕКТЫ И МЕТОДЫ ИССЛЕДОВАНИЙ

Исследования проводились в математической среде разработки `gblup.R2` и `blup.R3` в языке R либо программными модулями пакета SAS MIXED4 в операционной среде Unix.

### Предыстория развития вычисления уравнений линейных моделей

Исторические предпосылки и открытие моделей BLUE, а также ММЕ (уравнений смешанной модели, *mixed model equation*) стали результатом задач, которые поставили статистики [2]. Эта проблема была изящно решена с использованием наиболее вероятных методов производственных возможностей Дж. Лаша. Решение этой задачи привело к написанию дипломной работы Ч. Хендерсона, тема которой была посвящена попытке объединить теорию наименьших квадратов и селекционный индекс для оценки общих, специфических и материнских фенотипических способностей у свиней [3]. Впервые уравнения ММЕ были представлены в 1949 г. [4]. Однако доказательство того, что решения для неизменных (фиксированных) эффектов совпадают с решениями, полученными из обобщенных уравнений наименьших квадратов, появилось только в 1956 г. и не было опубликовано до 1959 г. [5]. Доказательство того, что решения для случайных (рандомизированных) эффектов такие же, как критерий селекционного индекса, наряду с обобщенными решениями по методу наименьших квадратов для фиксированных эффектов, используемыми в качестве средних значений, также получено в 1956 г., но не опубликовано до 1963 г. [6]. Термин про-

гноза BLUP, основанный на BLUE и математической среде решений ММЕ, был впервые опубликован в 1962 г. [7]. Эта история имела продолжение в 1973 г., когда Ч. Хендерсон опубликовал тезисы в материалах симпозиума памяти Дж. Лаша [8].

### Узкое место прогноза племенной ценности

Допустим, что основная линейная модель будет иметь следующий вид:

$$y = X\beta + Zu + \varepsilon \quad (1)$$

где  $y$  – вектор наблюдаемых фенотипических величин;

$X$  – известная фиксированная матрица  $n \times p$  с рангом  $r$ ;

$\beta$  – неизвестный вектор фиксированных (неизменных) эффектов;

$Z$  – матрица известных фиксированных эффектов;

$u$  и  $\varepsilon$  – ненаблюдаемые случайные (рандомизированные) векторные переменные (случайных неизвестных ошибок) с нулевыми средними значениями.

Тогда получим следующую дисперсно-ковариационную структуру:

$$V \cdot \begin{bmatrix} u \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}, \quad (2)$$

где матрицы  $G$  и  $R$  известны и несингулярны.

Селекционеры почти всегда предполагают, что линейная модель подходит для прогноза, и во многих ситуациях это верно. Однако нельзя игнорировать такую возможность, что нелинейные модели для прогноза в некоторых ситуациях могут быть лучше [9]. Нелинейная методология расчетов может оказаться сложной для реализации, и, следовательно, определение «лучшего» и мера того, насколько «лучше», будут определять, насколько легко вообще реализуются нелинейные методы в каждом конкретном случае.

Чтобы создать модель прогноза BLUP, нужно предсказать линейные функции  $\beta$  и  $u$ , скажем, в виде  $K'\beta + M'u$ , с помощью линейной функции  $y$ , т.е.  $F'y$ , так что математическое отношение  $F'y$  будет равно математическому ожиданию  $K'\beta + M'u$ , в котором дисперсии ошибки прогнозирования  $F'y - K'\beta - M'u$  сведены к минимуму. При этом считается, что  $K'\beta$  поддается оценке. Если производные функции от дисперсии матрицы ошибок предсказания, которые включают матрицу мно-

2. <https://rdrr.io/cran/pedigree/src/R/gblup.R>

3. <https://rdrr.io/cran/Phenotype/man/blup.html>

4. <https://bit.ly/3tKE95e>

жителей Лагранжа (для обеспечения беспристрастности такого анализа), приравниваются к нулю, то получится решение для функции  $F$ .

Предсказание предстанет в виде выражения:

$$F'y = K'\beta + M'u, \quad (3)$$

где векторы  $\beta$  и  $u$  – это решения системы уравнений ММЕ по Ч. Хендерсону:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \cdot \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \quad (4)$$

Общее матричное выражение здесь охватывает модель по отцам (sire model, SM, BLUP-SM (модели животного (индивидуума, animal model, AM, BLUP-AM) [10] и другие модели с несколькими селекционными признаками (multiple-trait models, MTM) [11–16]. Термин «модель индивидуума» впервые опубликован в 1988 г. [17], но, конечно же, в качестве модели описан ранее, в начале 1970-х гг. [18]. Первое практическое использование BLUP-AM было отмечено в 1990 г. в сравнении показателей быков-производителей и их дочерей Северо-Востока США. В сравнении участвовало приблизительно 1200 быков-производителей голштинской породы. Оценка была проведена при помощи нескольких компьютерных программ, написанных на FORTRAN, на компьютерах IBM 360 с объемом памяти 128 кб [19–20].

Уравнения ММЕ, указанные выше, формируются без необходимости знаний о распределении значений признака  $y$ . Для проверки гипотезы или оценки компонентов дисперсии по максимальному правдоподобию (REML, restricted (or residual, or reduced) maximum likelihood) предполагается, что измеряемые наблюдаемые фенотипические величины селекционного признака распределены нормально (распределение Гаусса-Лапласа). Тем не менее предполагается, что матрицы ковариации  $G$  и  $R$  представлены известными, невырожденными и положительно определенными значениями. Ранее были описаны методы, которые следует использовать для вычислений в случае, если матрицы  $G$  и  $R$  неизвестны или обе они сингулярны [8, 21–22].

Вычислительные аспекты построения уравнений ММЕ крайне любопытны, было введено понятие эквивалентных моделей, чтобы упростить вычисления или понимание отдельных моделей. Например, принято, что две разные модели эквивалентны, если они дают одно и то же математическое ожидание и ковариационную функцию  $y$  [23]. Примером может служить ограниченная модель инди-

видуума, которая может казаться удобной в случаях, когда от одного спаривания рождается много потомства, например, у свиней, или когда оставляются очень мало особей в качестве родителей для будущих поколений животных [24].

Также предложена модель, где учтены эффекты взаимодействия между быками-производителями и стадами для учета влияния окружающей среды среди полусибсов в одном стаде [11].

Описана такая модель BLUE, где значения  $cov(u, e)$  не равны нулю, и показано, что для таких моделей можно записать следующую эквивалентную модель, в которой ковариационная матрица между случайным вектором модели и остаточным вектором равна нулю.

#### Проблема включения предварительной (предполагаемой) информации

Еще одной проблемой в прогнозировании племенной ценности является предсказание  $K'\beta + M'u$  за счет  $F'y + S'\beta$ , где  $E \cdot (K'\beta + M'u) = E \cdot (F'y + S'\beta)$ , а диагональ при этом  $V(K'\beta + M'u - F'y - S'\beta)$  – минимизирована.

Вектор  $S'\beta$  является априорной несмещенной оценкой  $S'\beta$ , которая оценена независимо от функции  $y$  и  $V(\beta) = P$ , что сингулярным не является. Результат решения уравнений ММЕ для этой проблемы:

$$\begin{bmatrix} X'R^{-1}X + P^{-1} & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \cdot \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} X'R^{-1}y + P^{-1}\beta \\ Z'R^{-1}y \end{bmatrix} \quad (5)$$

Таким образом, могут быть учтены сделанные ранее оценки фиксированных эффектов, таких как различия между породами, различия в возрасте отела и различия по половому признаку. Следует обратить внимание на то, что если  $P^{-1}$  равно нулю, то решение для функции  $\beta$  полностью вытекает из исходных данных по продуктивности, а если элементы  $P^{-1}$  велики, то решения становятся похожими на предыдущие величины. Аналогичные выражения можно вывести в контексте теории Байеса, он удобен тем, что в нем нет необходимости различать фиксированные и случайные факторы среды [25].

Правые части уравнений ММЕ также содержат априорную информацию о значениях функции  $u$ . Если для векторов  $\beta$  и  $u = 0$  не существует априорной информации, то в уравнениях она содержится:

$$\begin{bmatrix} X'R^{-1}X + P^{-1} & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \cdot \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} X'R^{-1}y + P^{-1}\beta \\ Z'R^{-1}y + G^{-1}u \end{bmatrix} \quad (6)$$

Эти уравнения могут применяться для оценки животных внутри заданного стада.



Если для этих особей известна племенная ценность их предков по всем другим стадам, то данная информация также может быть включена в эти уравнения для заданного стада. Это отличается от изначального подхода Ч. Хендерсона, в котором теория селекционного индекса (как матрицы коэффициентов, так и правая часть уравнений) использовалась только при внутростадной оценке молочных коров [26, 27].

**Смещенный (искаженный) прогноз племенной ценности**

В 1987 г. была отмечена возможность того, что смещенные предсказания могут не иметь существенной значимости в моделировании процессов в животноводстве, т. е. может существовать такое несмещенное предсказание, которое имеет меньшую среднеквадратичную ошибку, чем смещенное. Эта возможность нашла отражение в системе следующих уравнений [15, 28, 29]:

$$\begin{bmatrix} X'_1 R^{-1} X_1 & X'_1 R^{-1} X_2 P & X'_1 R^{-1} Z \\ P X'_2 R^{-1} X_1 & P X'_2 R^{-1} X_2 P + P & P X'_2 R^{-1} Z \\ Z' R^{-1} X_1 & Z' R^{-1} X_2 P & Z' R^{-1} Z + G^{-1} \end{bmatrix} \cdot \begin{bmatrix} \hat{\beta}_1 \\ \hat{\alpha}_2 \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'_1 R^{-1} y \\ P X'_2 R^{-1} y \\ Z' R^{-1} y \end{bmatrix} \quad (7)$$

В этих выражениях матрица  $X$  разбита на  $(X_1, X_2)$ ,  $\beta' = (\beta'_1 \beta'_2)$  и  $P = \beta'_2 \beta'_2$ , так что  $\hat{\beta}_2 = P \hat{\alpha}_2$ , и  $\hat{\beta}_2$  является априорным значением  $\beta_2$ . Важно отметить, что решения для  $\hat{\beta}_2$  пропорциональны  $\hat{\beta}_2$ , вектору априорных значений, которые используются для формирования матрицы  $P$ , и это свойство было сочтено нежелательным [30]. Байесовские методы вычислений могут здесь оказаться предпочтительнее, чем использование сложных уравнений, и могут позволить получить более точные решения по прогнозу BLUP в смысле, что среднеквадратичная ошибка будет меньше [30]. Предвзятые (смещенные) оценки в прогнозе также могут быть сформированы простым добавлением константы к диагональным элементам для фиксированных факторов в модели уравнений ММЕ.

Если использовать смещенные оценки, то необходимо применять методы измерения систематической ошибки и вычисления среднеквадратичных ошибок [31]. Если можно было бы показать, что систематическая ошибка в оценке невелика, то тогда распространение необъективного прогноза генетических качеств особей оказалось приемлемым для быков-производителей. Однако это могло бы иметь юридические последствия.

**Матрицы родственных отношений (родства)**

Историческими предпосылками важного открытия, касающегося быстрого метода

вычисления обратной матрицы отношений исходя из списка отдельных родословных не инбредной популяции [32], было нахождение численных способов вычисления матриц родства.

Модель оценки быков-производителей Северо-Востока США была изменена в 1975 г. таким образом, чтобы включить в анализ отношения между быками на основе их предков по отцу и бабушке по материнской линии (модель grandsire, MGS), и хотя использование аддитивных генетических отношений в генетических оценках было шагом вперед, это привело к нескольким новым проблемам [17, 33, 34].

Первой возникшей проблемой было влияние родственных отношений на генетические групповые взаимодействия. Групповые различия за счет многих лет применения искусственного осеменения были значительно меньше по величине, чем это наблюдалось ранее в моделях без искусственного осеменения. Дискуссии доходили до того, что если бы имелась полная информация о родословной для каждого быка, то проще поколения «слить», и каждого такого быка отнести к одному и тому же «фиксированному» базовому поколению, – настолько эти групповые различия в США могли быть близкими к нулю.

Однако, поскольку отношения основывались только на основании информации об отце, групповые решения отражали различия в выборе метода селекции по материнской родословной [34]. Одновременно был решен вопрос учета инбридинга в обратной матрице отношений, который требовал хронологической упорядоченности данных о предках [35]. Был разработан быстрый метод нумерации отношений до их инверсии, который включал коэффициенты инбридинга. Этот алгоритм оказался очень полезным для мониторинга показателей инбридинга, особенно в небольших популяциях. Как только известен коэффициент инбридинга, за этим сразу следует вычисление правильных обратных элементов.

Скорость сходимости итерационных алгоритмов вычислений для получения решений уравнений смешанной модели при включении в расчет матриц отношений изменилась не к лучшему. Без матрицы взаимосвязей решение для быков обычно достигалось после 15 циклов итерации от начальных значений ценности, равных нулю. После ввода дополнительных недиагональных элементов в отцовской части ММЕ вследствие изменения обратной матрицы отношений, необходимое количество раундов итераций увеличилось.

Возросло и число итераций, необходимых для достижения сходимости с моделью индивидуума (BLUP-AM), поскольку вышло так, что диагональные элементы в ММЕ больше не доминировали вне диагоналей. В итоге пришлось изменить стратегии для решения ММЕ, и оригинальная модель Ч. Хендерсона уже не подходила. Тридиагмализация, технология разреженных матриц, повторные итерации исходных данных, коэффициенты релаксации, рекурсивное трансформирование обратной матрицы, APY – это далеко не все процедуры, которые приходится заводчикам использовать сегодня [36, 37].

#### Генетическая группировка животных по связям между ними

На практике самые большие проблемы при использовании аддитивных генетических отношений заключаются в *невозможности* проследить родословную *каждого* индивидуума до общего базового поколения и гарантировать то, что животные базовой популяции представляют не отобранную группу особей. Если основная популяция животных возникла в случае их селекции, то необходимо изменить обратную матрицу родства с учетом отбора животных [38, 39]. Решения системы уравнений ММЕ в случае игнорирования генетического базиса (референтной популяции животных) будут, конечно же, похожи на результаты без учета этого факта, но здесь стоит проявить осторожность при интерпретации.

Например, в таком случае рекомендуется использовать матрицу планирования эксперимента с пометками об изменении поколений, а отсутствие полных родословных у каждого из животных необходимо соотнести с базисом, его генетическими группами [40]. Например, можно использовать фантомных родителей, которых затем можно сгруппировать по году рождения их потомства (или другим параметрам), немного изменив уравнения ММЕ [17, 41]. Эта информация изменит и матрицу родственных отношений, где для каждого индивидуума потребуется количество неизвестных родителей –  $m$ , и тогда:

$$x = \frac{4}{m+2}. \quad (8)$$

Затем величина  $x$  добавляется к диагональному элементу индивидуума,  $x/2$  – к элементам родителей (или родительским группам), и  $x/4$  – к диагональным и недиагональным элементам родителей.

Решения для аддитивных генетических эффектов индивидуума в уравнениях ММЕ представляют собой оценку функций груп-

повых эффектов для животных плюс их собственные аддитивные генетические эффекты (как отклонения от группового среднего). Функция групповых эффектов, включенных в решение каждого индивидуума, зависит от частичных знаний о родстве. Если в группе у родителей отсутствуют данные по полу или дате рождения их потомков, и эти неизвестные коровы и быки сгруппированы отдельно, то это часто объединяется в единую группу в уравнениях, т. к. эти данные похожи, и в результате они оцениваются при помощи отдельной матрицы коэффициентов в уравнениях ММЕ.

Дать определение, какие функции групповых эффектов можно оценить, – непростая задача. Для того, чтобы уменьшить число фантомных групповых уравнений, предложено разбивать группы для неизвестных быков-производителей и коров по годам. Если задавать разные интервалы между поколениями для быка и коровы, то это тоже может устранить проблему. Животных, у которых неизвестны оба родителя, как отец, так и мать, тоже можно отнести к одной и той же фантомной группе. Это новый метод работы с отобранными животными генетического базиса, если их родословные неполные [39].

Было также показано, что в рамках концепции фантомных группировок напрямую меняется обратная матрица отношений [41]. Новое предложение концепции состоит в том, чтобы позволить каждому фантомному животному быть отдельной группой и иметь только одного потомка. Кроме того, тогда каждую особь можно проследить вплоть до ее генетического базиса, в котором все животные изначально считаются не отобранными (а фантомные животные вставляются в матрицу там, где это необходимо).

Чтобы учесть выбранных ранее фантомных животных после смены поколения «зеро» – «0», при формировании измененной инверсии необходимо использовать матрицу отношений. Каждое фантомное животное в этой матрице будет рассматриваться как «настоящее», а оба его родителя будут при этом фантомными. Таким образом, фантомные животные в этом предложении будут как бы случайными эффектами. Если поколение базиса удалено от текущих животных на 50 поколений, то файл родословной может содержать уже очень большое число фантомных животных. К счастью, нет необходимости строить инверсию в файле родословной для всех и «реальных» и фантомных животных, т. к. фантомные могут «поглощаться» в будущем

и перейти в «реальные». Это требует соблюдения правил для создания соответствующей «поглощенной», модифицированной обратной матрицы отношений исключительно для «реальных» животных. Однако при этом для

каждого «реального» индивидуума, для которого неизвестны один или оба родителя, должно быть известно число поколений, на сколько оно удалено от поколения «0».

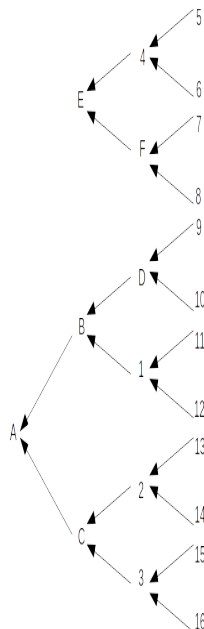


Рис. 1. Пример прослеживания родства животных относительно популяции «зеро» – «0»  
 Fig. 1. Example of tracing the kinship of animals relative to the “zero” - “0” population

Рассмотрим пример родословной (рис. 1), где фантомных животных обозначим цифрами, а «реальных» – буквами. Животные с 5-го по 16-е пусть образуют поколение «0». В табл. 1 приведена информация о каждом «реальном» животном, где  $M$  – число неизвестных родителей,  $t$  – число поколений, удаленных из базиса.

Таблица 1

Данные родства «реальных» животных  
 Relation data for “real” animals

Животное	Бык	Корова	m	t
A	0	0	0	3
B	0		1	2
C			2	2
D			2	1
E		0	1	2
F			2	1

Введем правила расчета для «поглощенных» животных, и модифицируем инверсию следующим образом:

– если величины  $m=0$ , тогда  $x=2$ , и  $x, \frac{-x}{2}, \frac{x}{4}$  добавляем к соответствующей позиции в инверсии, как указано ранее;

– если величины  $m=2$ , тогда  $x=1$ , вычисляем  $w = \frac{(2^t-1)}{(2^{t+1}-1)}$ , добавляем  $(x-w)$  к диагонали для этого индивидуа;

– если величины  $m=1$ , тогда  $x=4/3$ , вычисляем  $p = [x - \frac{(2^t-1-1)}{(2^t-1)}]^{-1}$  и добавляем  $(x - \frac{4p}{9})$  к диагонали индивидуа,  $(\frac{-x}{2} + 2\frac{p}{9})$  к диагонали «известного» родителя по диагонали и  $(\frac{x}{4} - \frac{p}{9})$  к известной диагонали «известного» родителя.

Результаты инверсии для «реальных» животных показаны в табл. 2.

Таблица 2

Результат инверсии для «реальных» животных  
 The result of the inversion for “real” animals

	A	B	C	D	E	F
A	2	-1	-1	0	0	0
B	-1	0	0	0	0	0
C	-1	0	0	0	0	0
D	0	0	0	0	0	0
E	0	0	0	0	0	0
F	0	0	0	0	0	8/9

Самая трудная часть этого метода заключается в определении изначально базисного поколения и количества поколений, после которых каждое животное будет удалено из базиса. Трудность вытекает из предположений о

средних интервалах между поколениями для быков-производителей, коров, матерей быков и матерей коров [41]. В свое время для североамериканской популяции молочного скота считалось, что подходящим базисным поколением могут быть животные, рожденные до 1945 г. Необходимо обратить внимание, что по мере увеличения количества поколений –  $t$ , удаленных от базиса,

$$w = \frac{(2^t - 1)}{(2^{t+1} - 1)}, \quad (9)$$

где  $w$  приближается к значению 0,5.

Таким образом, вместо того чтобы пытаться подсчитать количество поколений, стоит далее предположить, что любые новые животные, появившиеся в данных продуктивности в 1990 г., отошли от базиса на много поколений вперед. Необходимо дальнейшее изучение этого метода, т. к. могут возникнуть проблемы при решении уравнений путем итерации, поскольку, как показано в предыдущем примере, некоторые из диагональных элементов обратной матрицы отношений могут оказаться меньше единицы.

#### Неаддитивные генетические эффекты

Разработка новых моделей индивидуума для оценки неаддитивных генетических эффектов основана на выражении [42, 43]

$$y = X\beta + Zg + \epsilon, \quad (10)$$

здесь обозначения функций  $y$ ,  $Xp$ ,  $Z$  и  $\epsilon$  уже описаны ранее, а  $g$  – вектор суммарных генетических эффектов животных, который можно определить, например, как сумму аддитивных эффектов ( $a$ ), аддитивности аддукции ( $aa$ ), доминирования ( $d$ ), доминирования за счет доминирования ( $dd$ ), доминантный ( $ad$ ), аддитивности за счет действия доминантных генов ( $ad$ ), каждый из которых предполагается независимым от других эффектов. Ковариационная матрица  $g$  неинбредной популяции при этом находится в равновесии по сцеплению и равна матрице  $M$ , а именно:

$$M = V(a) + V(aa) + V(d) + V(dd) + V(ad) \quad (11)$$

И если  $A$  – матрица вычислений отношений, а  $D$  – доминантная матрица генетических отношений, тогда

$$M = A\sigma_a^2 + A \otimes A\sigma_{aa}^2 + D\sigma_d^2 + D \otimes D\sigma_{dd}^2 + A \otimes D\sigma_{ad}^2 \quad (12)$$

где  $\otimes$  – произведение по Адамару.  
Тогда

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + M^{-1} \end{bmatrix} \cdot \begin{bmatrix} \beta \\ g \end{bmatrix} = \begin{bmatrix} X^{-1}R^{-1}y \\ Z^{-1}R^{-1}y \end{bmatrix} \quad (13)$$

Далее, рассчитав решение для вектора  $g$ , получим

$$\hat{a} = AM^{-1}g; \hat{d} = DM^{-1}g; \quad (14)$$

$$aa = (A \otimes A) \cdot M^{-1}g; dd = (D \otimes D) \cdot M^{-1}g; ad = (A \otimes D) \cdot M^{-1}g \quad (15).$$

Проблема вычислений в этой модели заключается в сложности вычисления матрицы  $M^{-1}$ , так как классические правила вычисления матриц  $A^{-1}$  к  $M^{-1}$  неприменимы. Таким образом, потребуются сильная инверсия плеча матрицы  $M$  или необходимо построить эквивалентную систему уравнений ММЕ, которая не потребует инверсии матрицы. Были приняты также попытки оценить генетические эффекты доминирования и генетическую изменчивость доминирования на основании данных по продуктивности молочного скота, у которого путем переноса эмбрионов были созданы большие группы линейных потомков [22].

Обратное значение матрицы  $D$  можно вычислить, конвертируя  $D$  в диагональную матрицу, состоящую из блоков, которые достаточно малы для того, чтобы их можно было инвертировать напрямую. На данный момент нет простых и понятных правил, как проводить инверсию матрицы  $D$  по данным списка родословных, но существуют разработки моделей, основанных на матрице родства гамет, которая допускает инбридинг [44]. С другой стороны, разрабатываются все новые процедуры для прямого вычисления инверсии ( $A \otimes A$ ), где матрица  $A$  основана на отношениях быков и MGS [45]. В будущем, скорее всего, будут разработаны более эффективные алгоритмы для инвертирования ковариационных матриц для различных типов воздействия генов, до того момента, как у нас появятся подходящие для применения этих методов данные. Тем не менее предположение о равновесии сцепления должно быть упрощено при помощи соответствующих процедур, разработанных для вычисления корректных ковариационных матриц и основанных на способах инвертирования аддитивной генетической матрицы отношений.

#### Критерии выбора стратегий селекции

Важной областью будущих исследований для селекционеров-зоотехников является разработка новых процедур прогноза племенной ценности, которые должным образом учтут различные виды предвзятости при выборе стратегии селекции [11, 46, 47]. Отбор, по определению, – это любой процесс, есте-



ственный или искусственный, который препятствует выживанию или размножению одних особей или позволяет это делать другим. Ключом к пониманию предвзятости, вызванной отбором, является точное описание самого процесса, который может осуществляться по элементам функций  $y$ ,  $u$  и  $e$  или по их комбинациям.

Смысл выбора элементов функций  $u$  и  $e$  трудно осознать, потому что они являются ненаблюдаемыми. Отбор может быть как прямым, так и косвенным, через коррелированную переменную, скажем, например, функцию  $w$ , или комбинацию переменных. Отбор может быть описан линейными функциями, происходящими от  $y$ ,  $w$  или  $y$  и  $w$ , или нелинейными функциями. Таким образом, описание процесса выбора животных требует определения переменной или переменных, на основе которых принимаются решения, и правил или принципов принятия этих решений. Это непростая задача. Используя дефиниции, определяющие функцию  $w$  как вектор информации о всех кандидатах животных, которые доступны для отбора, такие, что функция  $w$  коррелирует с величинами функций  $y$ ,  $u$  или  $e$ , следующим образом:

$$E \begin{bmatrix} y \\ u \\ e \\ w \end{bmatrix} = \begin{bmatrix} X\beta \\ 0 \\ 0 \\ d \end{bmatrix} \quad (16)$$

$$V \begin{bmatrix} y \\ u \\ e \\ w \end{bmatrix} = \begin{bmatrix} V & ZG & R & C_{yw} \\ G & 0 & & C_w \end{bmatrix} \quad (17)$$

Матрицы  $G$ ,  $R$ ,  $C_{yw}$ ,  $C_{uw}$ ,  $C_{ew}$ , и  $C_w$  содержат известные значения параметров перед отбором, при этом решено описывать процессы отбора как функции выбраковки  $w$ , скажем,  $L'w$ . Разработан ряд различных матриц  $L'$  для решения конкретных задач, но явных указаний по формированию  $L'$  в общем контексте не существует. Большинство примеров матриц  $L'$  были такими, что строки матрицы  $L'$  добавляли к нулю, и, следовательно, каждая строка описывала как бы «разницу» в элементах  $w$  между выбранной особью (или группой особей) от тех особей, которые были выбракованы. Для иллюстрации предположим, что  $w$  – вектор длиной 7 и что первые три индивидуума предпочтительнее остальных четырех. Тогда матрица  $L'$ , которая отражала бы эти решения, выглядела бы так:

$$L' = \frac{1}{4} \cdot \begin{bmatrix} 4 & 0 & 0 & -1 & -1 & -1 & -1 \\ 0 & 4 & 0 & -1 & -1 & -1 & -1 \\ 0 & 0 & 4 & -1 & -1 & -1 & -1 \end{bmatrix} \quad (18)$$

Если бы три отобранных индивида были выбраны как группа, а не по отдельности, то матрица  $L'$  бы выглядела так:

$$L' = \frac{1}{12} \cdot [444 - 3 - 3 - 3 - 3] \quad (19)$$

Если бы функция  $w$  представляла отдельных коров, а анализ проводился через модель SM, то использование вышеуказанной выбраковки для учета селекции могло бы привести к очень схожим значениям прогноза племенной ценности для быков. Если, однако, анализ проводился бы на модели AM, то использование выбраковки по группам по сравнению с выбраковкой по отдельности могло бы привести к различным величинам прогноза племенной ценности коров, поскольку выбраковка учитывала бы не индивидуальные различия, а была бы скорректирована на среднюю разницу для группы из трех выбранных коров. Таким образом, правило отбора и модель анализа могут иметь решающее значение при устранении предвзятости отбора из прогноза племенной ценности.

В случае принятия модели отбора по Пирсону можно было бы применить многомерную нормальность для получения результатов, представленных в 1985 г. [46, 48], сделав заключения:

$$B = C_{yw}L; B_u = C_{uw}L; B_e = C_{ew}L; H = L'C_wL. \quad (20)$$

Если учитывать «условные» ожидания в контексте условно заданного отбора на величину функции  $w$ , тогда это отличалось бы от условно заданной вектором  $L'w$ . Термин «условный» не следует использовать в этом контексте. Матрицы ожидания и ковариации  $W$ , учитывающие выбор (обозначаются  $E_s$  и  $V_s$ ), равны:

$$E_s \cdot \begin{bmatrix} y \\ u \\ e \\ L'w \end{bmatrix} = \begin{bmatrix} X\beta \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} BH^{-1}\alpha \\ B_uH^{-1}\alpha \\ B_eH^{-1}\alpha \\ \alpha \end{bmatrix} \quad (21)$$

$$V_s \cdot \begin{bmatrix} u \\ e \\ L'w \end{bmatrix} \quad (22)$$

Здесь матрицы  $H_0 = H^{-1}(H - HS)H^{-1}$ , величина  $\alpha$  и матрица  $H_s$  – это средние значения и дисперсии заданной выборки особей. Как правило, значения  $\alpha$  и  $H_s$  получить трудно, не предполагая определенного типа отбора (например, отбор по усечению) и зная интенсивность отбора.

В животноводстве селекция усечением представляет стандартный метод отбора для разведения в следующем поколении. Животных ранжируют по их фенотипической ценности по некоторым признакам, например, по молочной продуктивности, и воспроизводят самый высокий процент. Эффекты

усеченного отбора для непрерывного селекционного признака могут быть смоделированы стандартным уравнением модели с использованием величин параметров наследуемости и усеченных нормальных распределений. Для бинарного признака можно легко смоделировать модель с выбором порога. Это считается простым и эффективным методом разведения. В количественной генетике это метод отбора, используемый в генетических алгоритмах для выбора потенциальных решений – кандидатов для рекомбинации, смоделированных после выбора метода разведения. При усеченном отборе решения кандидаты упорядочиваются по их пригодности, и некоторая их доля  $p$  (например,  $p=1/2, 1/3$  и т. д.) наиболее приспособленных особей отбирается и воспроизводится  $1/p$  раз. Выбор усечения менее сложен, чем другие методы селекции, и используется на практике не часто, а лишь в генетическом алгоритме селекции.

Уравнения смешанной модели [46], в которых учтены вышеуказанные ожидания, примут вид:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}B_e \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}B_e - G^{-1}B_u \\ B_e'R^{-1}X & B_e'R^{-1}Z - B_u'G^{-1} & B_e'R^{-1}B_e + B_u'G^{-1}B_u \end{bmatrix} \cdot \begin{bmatrix} \beta \\ \alpha \\ \epsilon \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ B_e'R^{-1}y \end{bmatrix} \quad (23).$$

Замечательной особенностью данной модели является то, что матрица  $H_s$  и величина  $\alpha$  в ней не проявляются и, следовательно, они не нужны для получения несмещенных прогнозов. Однако решения функции  $\beta$  даются вычислением функции  $H^{-1}\alpha$ .

Проблема селекции с моделью Пирсона заключается в природе предполагаемого процесса выбора особей [48]. Предположим, что есть две коровы (А и В) с записями продуктивности, и они ранжированы на основе базиса их записей как  $A > B$ . Если бы эту ситуацию симулировали в рамках модели Пирсона, то наблюдались бы только те примеры, в которых соблюдается условие  $A > B$ , и они были бы приняты в повторном исследовании выборки. Это отличает модель от таких схем выбора, где ранжирование  $A > B$ , а иногда  $B > A$ . Таким образом, здесь ранжирование животных является случайным, и, следовательно, соответствующие матрицы плана эксперимента в модели будут определяться тоже случайно [49, 50].

Последний процесс формирования выборки кажется более реалистичным, чем тот, который предполагается в модели Пирсона и используется в моделях Хендерсона. Но учитывая тот факт, что обычно имеется только одна единственная выборка данных, то про-

цесс их выбора, наблюдаемый при повторной выборке, тоже может быть предметом обсуждения других существующих сегодня альтернатив [46, 49, 50].

### Выбраковка коров

Рассмотрим случай с пятью коровами, прошедшими первую лактацию, из которых у двух была вторая. Трех коровам в одних условиях вводили лекарство bST (бычий соматотропин), а двум – нет (табл. 3). Модель без учета других воздействий окружающей среды будет следующей:

$$y_{ijk} = \mu + b_i + a_j + p_j + e_{ijk}, \quad (24)$$

где  $y_{ijk}$  – наблюдаемые значения удоя;  
 $\mu$  – современное групповое среднее (фиксированное);  
 $b_i$  – эффект от лечения bST (фиксированный);  
 $a_j$  – аддитивный генетический эффект животного (случайный, рандомизированный);  
 $p_j$  – фиксированный (постоянный) эффект коровы от окружающей среды (случайный);  
 $e_{ijk}$  – остаточный эффект (случайный) ошибки модели.

Пусть  $a, p$  и  $e$  – это векторы аддитивных генетических, постоянных экологических и остаточных эффектов соответственно, тогда:

$$V(a) = A\sigma_a^2 \quad (25)$$

$$\text{где } A = \frac{1}{16} \cdot \begin{bmatrix} 16 & 8 & 2 & 3 & 8 \\ 8 & 16 & 2 & 3 & 8 \\ 2 & 2 & 17 & 4 & 12 \\ 3 & 3 & 4 & 16 & 2 \\ 8 & 8 & 12 & 2 & 18 \end{bmatrix}$$

и  $\sigma_a^2$  являются аддитивной генетической дисперсией.

$$V(p) = I\sigma_p^2; V(e) = I\sigma_e^2, \quad (26)$$

где  $\sigma_p^2$  – постоянная дисперсия окружающей среды;

$\sigma_e^2$  – остаточная дисперсия ошибки.

Если наследуемость равна 0,25, а повторяемость 0,5, тогда

$$k_a = \frac{\sigma_a^2}{\sigma_a^2}; k_p = \frac{\sigma_p^2}{\sigma_p^2}. \quad (27)$$

Предположим, что отбор проводился в первые лактации, так что  $w=y$ , и отбор относится к типу  $L'y$ . Соответствующая матрица  $L'$  может быть такой:

$$L' = \frac{1}{4} \cdot \begin{bmatrix} -1 & 4 & -1 & -1 & -1 & 0 & 0 \\ -1 & -1 & -1 & 4 & -1 & 0 & 0 \end{bmatrix} \quad (28)$$

для  $y' = (5800 \ 6100 \ 5700 \ 6300 \ 5900 \ 6200 \ 6200)$

$$L'y = \begin{bmatrix} 175,0 \\ 425,0 \end{bmatrix} \quad (29)$$

Чтобы получить соответствующую систему решений ММЕ для типа отбора  $L'y$ , мы имеем:

$$B_s = Cov(s, L'y) = RL; B_u = Cov(u, L'y) = GZ'L \quad (30)$$

Тогда:

$$\begin{aligned} X'R^{-1}B_s &= X'L; Z'R^{-1}B_s - G^{-1}B_u = \\ &= 0; B_s'R^{-1}B_s + B_uG^{-1}B_u = L'(ZGZ' + R)L = L'VL \\ B_s'R^{-1}y &= L'y \end{aligned} \quad (31)$$

Уравнения разрешаются так:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}Z & X'L \\ Z'R^{-1}X & Z'R^{-1}Z + A^{-1}\lambda_a & Z'R^{-1}Z & 0 \\ Z'R^{-1}X & Z'R^{-1}Z & Z'R^{-1}Z + I\lambda_p & 0 \\ L'X & 0 & 0 & L'VL \end{bmatrix} \cdot \begin{bmatrix} \beta \\ d \\ p \\ t \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ Z'R^{-1}y \\ L'y \end{bmatrix}$$

где  $\lambda_a = \sigma_a^{-2}$  и  $\lambda_p = \sigma_p^{-2}$ .

Таблица 3

Пример данных для выбраковки коров  
Example of data for culling cows

Корова	Лечение заболевания	Удой, кг	
		лактация 1	лактация 2
1	Да	5800	
2	Да	6100	6200
3	Да	5700	
4	Нет	6300	6200
5	Нет	5900	

Решения для этого примера, наряду с решениями без учета отбора и решений с использованием матрицы  $L'y$ , для групп от-

бранных животных приведены в табл. 4. Решения, учитывающие такой отбор, не приводят к каким-либо изменениям в ранжировании, лишь незначительно отличаются по абсолютной величине прогноза от решений, не учитывающих тип отбора. Решения для функции  $t$ , по которым вычисляют  $H^{-1}\alpha$ , представляют собой модификации матрицы  $L'y$ , скорректированные с учетом различий в группах с лечением bST по сравнению с группами без лечения, но с учетом родства между коровами. Решения с двумя методами учета типа отбора меньше отличаются друг от друга, чем от решений, целиком игнорирующих селекцию.

Симуляция с соответствующим процессом выбора данных является необходимой для определения того, какой метод учета выбраковки коров следует использовать. В этом примере матрицы  $L'X$  не были нулевыми, поскольку одно животное было выбрано из группы, получавшей терапию, и одно животное – из необработанной группы. Матрица  $L'X$  также не была бы нулевой, если бы в модель были включены ковариаты. Также выбраковка коров могла быть полностью проигнорирована, если бы  $L'X$  была нулевой, например, если бы отбор полностью находился в пределах подклассов с фиксированными эффектами.

Таблица 4

Решения прогноза выбраковки коров с учетом двух методов отбора (индивидуальный и групповой) и без учета отбора  
Decisions of culling forecast with consideration of two selection methods (individual and group) and without regard for selection

Решение для метода	Метод селекции		
	Без учета метода, кг	Индивидуальный базис, кг	Групповой базис, кг
bST и без	-226,92	-142,94	-159,21
<i>Аддитивное генетическое окружение</i>			
Корова 1	-12,27	-18,09	-16,96
Корова 2	30,42	30,43	30,43
Корова 3	-62,04	-64,86	-64,32
Корова 4	41,34	51,54	49,56
Корова 5	-59,73	-60,34	-60,22
<i>Неизменное окружение</i>			
Корова 1	-23,07	-35,96	-33,46
Корова 2	62,32	61,07	61,31
Корова 3	-39,82	-53,70	-51,01
Корова 4	50,13	64,68	61,94
Корова 5	-49,56	-36,19	-38,73
$t_1$		298,96	468,15
$t_2$		328,34	

**Ассоциации между стадами и быками-производителями**

Многие модели прогноза племенной ценности быков-производителей учитывали влияние эффектов «стадо – год – сезон» (HYS), которые логично считать случайной величиной. Из-за возможности того, что лучшие быки связаны с лучшим стадом коров, эффекты HYS рассматривались в уравнениях ММЕ как фиксированные эффекты. Показано, что ожидания от быков были равны ожиданиям истинной ценности быков в соответствии с заданной моделью отбора.

Для стран с небольшим размером стад существует множество подклассов HYS, где участвуют только 3 или менее телок в первой лактации и фиксированные эффекты HYS. В большинство подклассов в таких случаях не будет вноситься никакая информация в уравнения для быка вследствие поглощения уравнений HYS. Предполагается, что в таких странах связь между быками не существует и просто принято считать эффекты HYS случайными величинами. Если же связь действительно существует, то эффекты HYS также можно считать случайными, но селекционеру придется признать, что исходные решения по прогнозу не будут полностью объективными и, возможно, не будут иметь меньшую среднеквадратичную ошибку.

Пример показывает, что модели, разработанные для одной страны (и ситуации) не могут быть непосредственно перенесены в другие страны (и ситуации), универсальной модели для всех стран или пород не существует. Была определена связь между быками и эффектами HYS как связь между истинной племенной ценностью производителей и истинными эффектами HYS, причем оба из которых не наблюдаемы. В определении линейной модели это будет матрица  $L'u$  типа отбора, где функция  $u$  будет вектором истинных эффектов HYS:

$$y = Xb + Za + Fh + e, \quad (33)$$

где  $y$  – вектор наблюдаемой продуктивности;

$b$  – вектор фиксированных эффектов;

$a$  – случайный вектор аддитивных генетических эффектов индивидуума (быка-производителя);

$h$  – случайный вектор эффектов HYS;

$e$  – случайные  $y = Xb + Za + Fh + e$ ных эффектов;

$X, Z$  и  $F$  – соответствующие матрицы инцидентности (бинарного отношения между двумя различными типами объектов).

В ситуации отсутствия отбора

$$V(a) = A\sigma_a^2; V(h) = I\sigma_h^2; V(e) = I\sigma_e^2 \quad (34)$$

Если  $w = h$ , то матрица  $L'$  может состоять из  $s$  строк для  $s$  быков с потомством в этих HYS-подклассах, и каждая строка в матрице будет отражать «различие» в величинах истинных эффектов HYS тех подклассов, в которых каждый бык имел потомство от тех подклассов, в которых самец не имел потомства. Каждый бык-производитель должен иметь по крайней мере один подкласс, в котором у него нет потомства, чтобы вообще было сформировано определение матрицы  $L'$ . Вероятно, будут такие быки, которые в прошлом имели потомков, которые принадлежат на самом деле к одним и тем же подклассам, тогда для них две строки в  $L'$  будут идентичны. Точно так же некие два быка могут иметь потомство в точно противоположных подклассах, тогда для них два ряда в  $L'$  будут одинаковыми, за исключением знака. В  $L'$  в итоге должна быть включена лишь одна строка. Если бы все быки имели потомство во всех подклассах, тогда не было бы необходимости учитывать неслучайную связь производителей с HYS.

Для выбора типа  $L'u$  и модели, приведенной выше, рекомендуется использовать следующие уравнения:

$$\begin{bmatrix} X'X & X'Z & X'F & 0 \\ Z'X & Z'Z + A^{-1}k_s & Z'F & 0 \\ F'X & F'Z & F'F + Ik_a & -L \\ 0 & 0 & -L' & L'Lk_h^{-1} \end{bmatrix} \cdot \begin{bmatrix} \beta \\ a \\ h \\ e \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ F'y \\ 0 \end{bmatrix} \quad (35).$$

Рассмотрим в качестве примера данные (табл. 5) по коровам 5 вариантов HYS от 4 быков. Эти данные были проанализированы на 5 моделях  $y = Xb + Za + Fh + e$  для некоторых эффектов приведены в табл. 5. Одна из моделей игнорировала быка в ассоциации с HYS, и эффекты HYS в анализе были случайными. Во второй модели эффекты HYS рассматривались как фиксированные, а в третьей модели использовались уравнения, где матрица  $L'$  формировалась как

$$L' = \frac{1}{6} \cdot \begin{bmatrix} 2 & 2 & 2 & -3 & -3 \\ 2 & 2 & -3 & 2 & -3 \\ -2 & -2 & -2 & 3 & 3 \\ -3 & 2 & 2 & -3 & 2 \end{bmatrix} \quad (36)$$

с одним рядом на каждого быка. В этом примере строки для быков 1 и 3 совпадают, за исключением знака, поэтому строка 3 была для этой модели при анализе исключена. Последние две модели описывают связь меж-



ду генотипами животных ( $a$ ) и эффектами HYS ( $h$ ) посредством ковариации:

$$V \cdot \begin{bmatrix} h \\ a \end{bmatrix} = \begin{bmatrix} I\sigma_h^2 & S\sigma_{ah} \\ S'\sigma_{ah} & A\sigma_a^2 \end{bmatrix} \quad (37)$$

Предположим, что для всех моделей

$$\begin{aligned} \sigma_a^2 &= 0,25\sigma_y^2; \sigma_h^2 = 0,375\sigma_y^2; \\ \text{и } \sigma_{ah} &= 0,03\sigma_y^2, \text{ и } \sigma_{ah} = 0,20\sigma_y^2 \end{aligned} \quad (38)$$

Матрица  $S$  будет эквивалентна  $F'Z$ . Аддитивные генетические племенные ценности коров, записанных в HYS, коррелируют с этими эффектами. Первая модель получилась аналогичной двум последним моделям, за ис-

ключением того, что ковариация между  $a$  и  $h$  считалась равной 0. Решения для 5 моделей оценки эффектов HYS и быка (выраженные как отклонения от первого эффекта HYS и первого быка соответственно) значительно различались. Поскольку ковариация между  $a$  и  $h$  увеличилась, то диапазон различий по быкам и эффектам HYS уменьшился, однако рейтинг быков и эффектов HYS при этом не изменился. В модели с эффектами HYS, выбранными фиксированными, так и с использованием уравнения (19) привело к переоценке быков 2 и 3 в первой модели, что и в данном случае привело к увеличению различий в эффектах HYS.

Таблица 5

Пример данных для производителей по проблеме «стадо – год – сезон» (HYS)  
Sample data for producers on the herd - year - season (HYS) problem

HYS	Номер быка	Номер коровы	Удой по 1-й лактации, кг
1	1	5	6100
1	1	6	6300
1	2	7	5800
2	1	8	6500
2	2	9	6000
2	4	10	5000
3	1	11	6400
3	1	12	6200
3	4	13	5200
4	2	14	5900
4	2	15	5700
4	3	16	5600
5	3	17	5500
5	3	18	5800
5	4	19	5400
5	4	20	5100

В случае выбора – какая из моделей наиболее подходит для этого решения – необходимо исходить из того, какая модель лучше всего описывает процесс формирования выборки эксперимента. Модель ковариации допускает среднюю степень ассоциации для всех типов HYS и животных. На практике степень ассоциации генотипов животных с эффектами HYS может различаться для разных стад и может не иметь линейной зависимости. Использование ковариационной модели требует инверсии (33), что исключает использование быстрого метода для инвертирования матрицы отношений.

Рассмотрение эффектов HYS как фиксированных объясняет систематическую ошибку (смещение) в генетических оценках животных, но этот прогноз не обязательно будет

лучшим. С другой стороны, формирование матрицы  $L'$  и решение (31) может быть очень затратным. Степень смещения, возникающего в генетической оценке, если эта ассоциация игнорируется, не была количественно определена.

**Оценка смещения анализа при выборе животных**

Естественным следствием высоких цен на сперму и эмбрионы генетически более совершенных особей является то, что часто наблюдается предвзятость к потомству таких животных (PREF, animal preferences), т. е. потомство обеспечивается более благоприятной средой при кормлении и содержании, чем потомство животных среднего или ниже среднего уровня ценности. Выявление и учет PREF-животных невозможны без ежедневного наблюдения за

ними в каждом стаде. Когда PREF были идентифицированы впервые, было предложено удалить их записи из базы данных. Некоторые PREF можно рассматривать как результат отбора по остаточным эффектам, например, в линейной модели, или, по крайней мере, в качестве остатка, в котором бы такие животные получали измененную селекционером оценку [46].

Существует простой подход к решению проблемы, заключающийся в добавлении в модель групп с PREF и назначении коров в эти группы. Простым способом группировки может быть цена спермы их быков-производителей на момент зачатия этой особи. Все потомство одного быка не будет (и не должно быть) отнесено к одной и той же группе PREF. Обычно цена спермы быка меняется по мере того, как бык переходит от начальной фазы выборки к своей первой оценке племенной ценности, и к оценке, основанной на дочерях второго цикла осеменения. Для реализации этой модели записи о ценах на сперму быков необходимо накапливать, а затем объединять с данными о регистрации данных по продуктивности их дочерей в соответствии с датой отела. Это очень сложная задача, требующая сотрудничества всех станций искусственного осеменения и РИСЦ, т. к. ведение и существование полных прайс-листов на всех быков в настоящее время сомнительно. Анализ на модели АМ для PREF-групп (очень похоже на пример с эффектом групп лечения коров) может предоставить вычисление различий от среднего при лечении на коров в каждой группе с поправкой на все другие эффекты, используемые в модели. Затем эти различия могут быть проверены на значимость и достоверность. Кроме того, коровы, которые подвергались предпочтительному лечению, будут приспособлены для этого лечения, и это может быть эффектом группы PREF, к которой они были отнесены. Тот факт, что мы не знаем, с какими коровами обращались более предпочтительно, не поможет нам определить, какие коровы все еще переоценены.

Создание больших полнородственных групп коров путем эмбриотрансфера может потребовать формирования групп PREF на основе ценности обоих родителей (быка-производителя и матери-донора), а не только быка. Если недоступна информация о ценах на сперму, то для оценки производственных селекционных признаков возможна группи-

ровка коров на основе оценки типа их отцов и матерей. Простая стратегия заключалась бы в использовании типов классификации самой коровы. Однако если существует генетическая связь между типом и продуктивностью, то стратегия использования типовых данных для быков-производителей, матерей или коровы не может применяться. Предвзятое обращение можно также рассматривать как связь между аддитивными генетическими эффектами животных и остаточными эффектами, с которыми можно справиться вышеописанными алгоритмами оценки производственных селекционных признаков.

## ВЫВОДЫ

1. Детали генетической оценки животных описаны в монографии [15], и в настоящее время все исследования в области селекции и воспроизводства сельскохозяйственных животных основаны на этих алгоритмах. При этом вопрос о непредвзятости оценки племенной ценности остается открытым. Важным открытием являлись способы инвертирования матрицы аддитивных генетических отношений. Это привело к применению моделей индивидуума (АМ) для племенной генетической оценки и исследованию неаддитивных генетических эффектов. Учет инбридинга и отбора животных генетического базиса, а также уместность фиктивных родительских групп в отношениях решается при помощи матриц генетических и неаддитивных генетических отношений.

2. Данная область исследований при становлении племенной оценки при помощи BLUP в России будет очень значимой, но в техническом исполнении сложной и трудной. Вычисление генетических параметров и племенной ценности по данным о субъектах также вызовет сложности в методологии [4]. Ключевым моментом будет точное описание методологии выбора животных и данных по ним, правил выбора и предполагаемых моделей селекции. Основной вклад, охватывающий широту тем и оказавший большое влияние на все области математической статистики в животноводстве, стали результатом усилий одного человека, Чарльза Хендерсона, и мировая прикладная наука о животноводстве навсегда останется у него за это в долгу.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Lush J.L.* The bull index problem in the light of modern genetics // *Journal of Dairy Science*. – 1933. – Vol. 16, N 6. – P. 501–522. – [https://doi.org/10.3168/jds.S0022-0302\(33\)93369-X](https://doi.org/10.3168/jds.S0022-0302(33)93369-X).
2. *Mood A.M.F.* Introduction to the Theory of Statistics. – 1950. – P. 1–17.
3. *Henderson C.R.* Estimation of general, specific and maternal combining abilities in crosses among inbred lines of swine. – Iowa State University, 1948. – P. 25–27.
4. *Henderson C.R.* Estimation of changes in herd environment // *J. Dairy Sci.* – 1949. – Vol. 32, N 8. – P. 706–706.
5. *Misztal I., Stein Y., Lourenco D.A.L.* Genomic evaluation with multibreed and crossbred data // *JDS Communications*. – 2022. – P. 1-10. – <https://doi.org/10.3168/jdsc.2021-0177>.
6. *Barwick S.A., Henzell A.L.* Development successes and issues for the future in deriving and applying selection indexes for beef breeding // *Australian Journal of Experimental Agriculture*. – 2005. – Vol. 45, N 8. – P. 923–933. – <https://doi.org/10.1071/EA05068>.
7. *Misztal I., Aggrey S.E., Muir W.M.* Experiences with a single-step genome evaluation // *Poultry science*. – 2013. – Vol. 92, N 9. – P. 2530–2534. – <https://doi.org/10.3382/ps.2012-02739>.
8. *Shook G.E.* Major advances in determining appropriate selection goals // *Journal of Dairy Science*. – 2006. – Vol. 89, N 4. – P. 1349–1361. – [https://doi.org/10.3168/jds.S0022-0302\(06\)72202-0](https://doi.org/10.3168/jds.S0022-0302(06)72202-0).
9. *Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses / D.A.L. Lourenco, I. Misztal, S. Tsuruta [et al.] // Journal of dairy science*. – 2014. – Vol. 97, N 3. – P. 1742–1752. <https://doi.org/10.3168/jds.2013-6916>.
10. *Genetic parameters and prediction of genotypic values for root quality traits in cassava using REML/BLUP / E.J. Oliveira, F.A. Santana, L.A. Oliveira [et al.] // Genetics and Molecular Research*. – 2014. – Vol. 13, N 3. – P. 6683–6700. – <http://dx.doi.org/10.4238/2014.August.28.13>.
11. *Genomic evaluation of cattle in a multi-breed context / M.S. Lund, G. Su, L. Janss [et al.] // Livestock Science*. – 2014. – Vol. 166. – P. 101–110. – <https://doi.org/10.1016/j.livsci.2014.05.008>.
12. *Henderson C.R.* General flexibility of linear model techniques for sire evaluation // *Journal of Dairy Science*. – 1974. – Vol. 57, N 8. – P. 963–972. – [https://doi.org/10.3168/jds.S0022-0302\(74\)84993-3](https://doi.org/10.3168/jds.S0022-0302(74)84993-3).
13. *Henderson C.R.* Comparison of alternative sire evaluation methods // *Journal of Animal Science*. – 1975. – Vol. 41, N 3. – P. 760–770. – <https://doi.org/10.2527/jas1975.413760x>.
14. *Henderson C.R.* Best linear unbiased prediction of breeding values not in the model for records // *Journal of Dairy Science*. – 1977. – Vol. 60, N 5. – P. 783–787. – [https://doi.org/10.3168/jds.S0022-0302\(77\)83935-0](https://doi.org/10.3168/jds.S0022-0302(77)83935-0).
15. *Schaeffer L.R.* Application of random regression models in animal breeding // *Livestock Production Science*. – 2004. – Vol. 86, N 1–3. – P. 35–45. – [https://doi.org/10.1016/S0301-6226\(03\)00151-9](https://doi.org/10.1016/S0301-6226(03)00151-9).
16. *Henderson C.R., Quaas R.L.* Multiple trait evaluation using relatives' records // *Journal of Animal Science*. – 1976. – Vol. 43, N 6. – P. 1188–1197. – <https://doi.org/10.2527/jas1976.4361188x>.
17. *Quaas R.L.* Additive genetic model with groups and relationships // *Journal of Dairy Science*. – 1988. – Vol. 71, N 5. – P. 1338–1345. – [https://doi.org/10.3168/jds.S0022-0302\(88\)79691-5](https://doi.org/10.3168/jds.S0022-0302(88)79691-5).
18. *Generalized estimation of the BLUP in mixed-effects models: A comparison with ML and REML / C.R. Yu, K.H. Zou, M.O. Carlsson, S. Weerahandi // Communications in Statistics-Simulation and Computation*. – 2015. – Vol. 44, N 3. – P. 694–704. – <https://doi.org/10.1080/03610918.2013.790445>.
19. *Schaeffer L.R.* CR Henderson: Contributions to predicting genetic merit // *Journal of dairy science*. – 1991. – Vol. 74, N 11. – P. 4052–4066. – [https://doi.org/10.3168/jds.S0022-0302\(91\)78601-3](https://doi.org/10.3168/jds.S0022-0302(91)78601-3).
20. *Calus M.P.L., De Haas Y., Veerkamp R.F.* Combining cow and bull reference populations to increase accuracy of genomic prediction and genome-wide association studies // *Journal of Dairy Science*. – 2013. – Vol. 96, N 10. – P. 6703–6715. – <https://doi.org/10.3168/jds.2012-6013>.
21. *Gianola D., Foulley J.L., Fernando R.L.* Prediction of breeding values when variances are not known // *Génétique sélection évolution*. – 1986. – Vol. 18, N 4. – P. 485–498.

22. Yang H., Su G. Impact of phenotypic information of previous generations and depth of pedigree on estimates of genetic parameters and breeding values // *Livestock Science*. – 2016. – Vol. 187. – P. 61–67. – <https://doi.org/10.1016/j.livsci.2016.03.001>.
23. Henderson C.R. Equivalent linear models to reduce computations // *Journal of Dairy Science*. – 1985. – Vol. 68, N 9. – P. 2267–2277. – [https://doi.org/10.3168/jds.S0022-0302\(85\)81099-7](https://doi.org/10.3168/jds.S0022-0302(85)81099-7).
24. Quaas R.L., Pollak E.J. Mixed model methodology for farm and ranch beef cattle testing programs // *Journal of Animal Science*. – 1980. – Vol. 51, N 6. – P. 1277–1287. – <https://doi.org/10.2527/jas1981.5161277x>.
25. Gianola D., Fernando R.L. Bayesian methods in animal breeding theory // *Journal of Animal Science*. – 1986. – Vol. 63, N 1. – P. 217–244. – <https://doi.org/10.2527/jas1986.631217x>.
26. Henderson C.R. Use of all relatives in intraherd prediction of breeding values and producing abilities // *Journal of Dairy Science*. – 1975. – Vol. 58, N 12. – P. 1910–1916. – [https://doi.org/10.3168/jds.S0022-0302\(75\)84808-9](https://doi.org/10.3168/jds.S0022-0302(75)84808-9).
27. Van Vleck L.D. Derivation of Henderson's method of incorporating artificial insemination sire evaluations into intraherd prediction of breeding values // *Journal of Dairy Science*. – 1982. – Vol. 65, N 2. – P. 284–286. – [https://doi.org/10.3168/jds.S0022-0302\(82\)82190-5](https://doi.org/10.3168/jds.S0022-0302(82)82190-5).
28. Chen X., Liang H., Wang Y. Total positivity of recursive matrices // *Linear Algebra and its Applications*. – 2015. – Vol. 471. – P. 383–393. – <https://doi.org/10.1016/j.laa.2015.01.009>.
29. Misztal I., Legarra A. Invited review: efficient computation strategies in genomic selection // *Animal*. – 2017. – Vol. 11, N 5. – P. 731–736. – DOI:10.1017/S1751731116002366.
30. Misztal I., Aggrey S.E., Muir W.M. Experiences with a single-step genome evaluation // *Poultry science*. – 2013. – Vol. 92, N 9. – P. 2530–2534. – <https://doi.org/10.3382/ps.2012-02739>.
31. Henderson C.R. Simple method to compute biases and mean squared errors of linear estimators and predictors in a selection model assuming multivariate normality // *Journal of Dairy Science*. – 1988. – Vol. 71, N 11. – P. 3135–3142. – [https://doi.org/10.3168/jds.S0022-0302\(88\)79914-2](https://doi.org/10.3168/jds.S0022-0302(88)79914-2).
32. Henderson C.R. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values // *Biometrics*. – 1976. – P. 69–83. – <https://doi.org/10.2307/2529339>.
33. Darvasi A. Experimental strategies for the genetic dissection of complex traits in animal models // *Nature genetics*. – 1998. – Vol. 18, N 1. – P. 19–24.
34. Pollak E.J., Quaas R.L. Definition of group effects in sire evaluation models // *Journal of Dairy Science*. – 1983. – Vol. 66, N 7. – P. 1503–1509. – [https://doi.org/10.3168/jds.S0022-0302\(83\)81965-1](https://doi.org/10.3168/jds.S0022-0302(83)81965-1).
35. Quaas R.L. Computing the diagonal elements and inverse of a large numerator relationship matrix // *Biometrics*. – 1976. – P. 949–953. – <https://doi.org/10.2307/2529279>.
36. Misztal I., Legarra A., Aguilar I. Using recursion to compute the inverse of the genomic relationship matrix // *Journal of dairy science*. – 2014. – Vol. 97, N 6. – P. 3943–3952. – <https://doi.org/10.3168/jds.2013-7752>.
37. Misztal I. Restricted maximum likelihood estimation of variance components in animal model using sparse matrix inversion and a supercomputer // *Journal of Dairy Science*. – 1990. – Vol. 73, N 1. – P. 163–172. – [https://doi.org/10.3168/jds.S0022-0302\(90\)78660-2](https://doi.org/10.3168/jds.S0022-0302(90)78660-2).
38. Henderson C.R. Best linear unbiased prediction using relationship matrices derived from selected base populations // *Journal of Dairy Science*. – 1985. – Vol. 68, N 2. – P. 443–448. – [https://doi.org/10.3168/jds.S0022-0302\(85\)80843-2](https://doi.org/10.3168/jds.S0022-0302(85)80843-2).
39. Assessing the contribution of breeds to genetic diversity in conservation schemes / H. Eding, R.P. Crooijmans, M.A. Groenen [et al.] // *Genetics Selection Evolution*. – 2002. – Vol. 34, N 5. – P. 613–633. – <https://doi.org/10.1051/gse:2002027>.
40. Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses / D.A.L. Lourenco, I. Misztal, S. Tsuruta [et al.] // *Journal of dairy science*. – 2014. – Vol. 97, N 3. – P. 1742–1752. – <https://doi.org/10.3168/jds.2013-6916>.
41. Genomic-polygenic and polygenic predictions for milk yield, fat yield, and age at first calving in Thai multibreed dairy population using genic and functional sets of genotypes / T. Laodim, M.A.



- Elzo, S. Koonawootrittriron [et al.] // *Livestock Science*. – 2019. – Vol. 219. – P. 17–24. – <https://doi.org/10.1016/j.livsci.2018.11.008>.
42. Zhou J. *Advances in Pedigree Analysis: Hardy-Weinberg Equilibrium, Strain Imputation, and Maternal Effects*. – University of California, Los Angeles, 2011.
43. Ahlborn-Breier G., Hohenboken W.D. Additive and nonadditive genetic effects on milk production in dairy cattle: evidence for major individual heterosis // *Journal of Dairy science*. – 1991. – Vol. 74, N 2. – P. 592–602. – [https://doi.org/10.3168/jds.S0022-0302\(91\)78206-4](https://doi.org/10.3168/jds.S0022-0302(91)78206-4).
44. Smith S.P., Mäki-Tanila A. Genotypic covariance matrices and their inverses for models allowing dominance and inbreeding // *Genetics Selection Evolution*. – 1990. – Vol. 22, N 1. – P. 65–91.
45. VanRaden P.M., Hoeschele I. Rapid inversion of additive by additive relationship matrices by including sire-dam combination effects // *Journal of dairy science*. – 1991. – Vol. 74, N 2. – P. 570–579. – [https://doi.org/10.3168/jds.S0022-0302\(91\)78204-0](https://doi.org/10.3168/jds.S0022-0302(91)78204-0).
46. Schaeffer L.R. CR Henderson: Contributions to predicting genetic merit // *Journal of Dairy Science*. – 1991. – Vol. 74, N 11. – P. 4052–4066. – [https://doi.org/10.3168/jds.S0022-0302\(91\)78601-3](https://doi.org/10.3168/jds.S0022-0302(91)78601-3).
47. Kahneman D., Thaler R.H. Anomalies: Utility maximization and experienced utility // *Journal of economic perspectives*. – 2006. – Vol. 20, N 1. – P. 221–234. – DOI: 10.1257/089533006776526076.
48. Single-step genomic evaluation using multitrait random regression model and test-day data / M. Koivula, I. Strandén, J. Pösö [et al.] // *Journal of Dairy Science*. – 2015. – Vol. 98, N 4. – P. 2775–2784. – <https://doi.org/10.3168/jds.2014-8975>.
49. Likelihood estimation of quantitative genetic parameters when selection occurs: models and problems / D. Gianola, R.L. Fernando, S. Im [et al.] // *Genome*. – 1989. – Vol. 31, N 2. – P. 768–777. – <https://doi.org/10.1139/g89-136>.
50. Gianola D., Im S., Fernando R.L. Prediction of breeding value under Henderson's selection model: a revisit // *Journal of dairy science*. – 1988. – Vol. 71, N 10. – P. 2790–2798. – [https://doi.org/10.3168/jds.S0022-0302\(88\)79873-2](https://doi.org/10.3168/jds.S0022-0302(88)79873-2).
51. Single Step, a general approach for genomic selection / A. Legarra, O.F. Christensen, I. Aguilar [et al.] // *Livestock Science*. – 2014. – Vol. 166. – P. 54–65. – <https://doi.org/10.1016/j.livsci.2014.04.029>.

## REFERENCES

- Lush J.L., The bull index problem in the light of modern genetics, *Journal of Dairy Science*, 1933, Vol. 16, No. 6, P. 501–522, [https://doi.org/10.3168/jds.S0022-0302\(33\)93369-X](https://doi.org/10.3168/jds.S0022-0302(33)93369-X).
- Mood A.M.F., *Introduction to the Theory of Statistics*, 1950, pp. 1–17.
- Henderson C.R., *Estimation of general, specific and maternal combining abilities in crosses among inbred lines of swine*, Iowa State University, 1948, pp. 25–27.
- Henderson C.R. Estimation of changes in herd environment, *J. Dairy Sci*, 1949, Vol. 32, No. 8, pp. 706–706.
- Misztal I., Stein Y., Lourenco D.A.L., Genomic evaluation with multibreed and crossbred data, *JDS Communications*, 2022, pp. 1–10, <https://doi.org/10.3168/jdsc.2021-0177>.
- Barwick S.A., Henzell A.L., Development successes and issues for the future in deriving and applying selection indexes for beef breeding, *Australian Journal of Experimental Agriculture*, 2005, Vol. 45, No. 8, pp. 923–933, <https://doi.org/10.1071/EA05068>.
- Misztal I., Aggrey S.E., Muir W.M., Experiences with a single-step genome evaluation, *Poultry science*, 2013, Vol. 92, No. 9, pp. 2530–2534, <https://doi.org/10.3382/ps.2012-02739>.
- Shook G.E., Major advances in determining appropriate selection goals, *Journal of dairy science*, 2006, Vol. 89, No. 4, pp. 1349–1361, [https://doi.org/10.3168/jds.S0022-0302\(06\)72202-0](https://doi.org/10.3168/jds.S0022-0302(06)72202-0).
- Lourenco D.A.L., Misztal I., Tsuruta S., Aguilar I., Ezra E., Ron M., Weller J.I., Lourenco D.A.L. [et al.], Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses, *Journal of dairy science*, 2014, Vol. 97, No. 3, pp. 1742–1752, <https://doi.org/10.3168/jds.2013-6916>.
- Oliveira E.J., Santana F.A., Oliveira L.A., Santos V.S. Oliveira E.J. [et al.], Genetic parameters and prediction of genotypic values for root quality traits in cassava using REML/BLUP, *Genetics*

- and Molecular Research*, 2014, Vol. 13, No. 3, pp. 6683–6700, <http://dx.doi.org/10.4238/2014.August.28.13>.
11. Lund M.S., Su G., Janss L., Guldbbrandtsen B., Brøndum R.F., Lund M.S. [et al.] Genomic evaluation of cattle in a multi-breed context, *Livestock Science*, 2014, Vol. 166, pp. 101–110, <https://doi.org/10.1016/j.livsci.2014.05.008>.
  12. Henderson C.R., General flexibility of linear model techniques for sire evaluation, *Journal of Dairy Science*, 1974, Vol. 57, No. 8, pp. 963–972, [https://doi.org/10.3168/jds.S0022-0302\(74\)84993-3](https://doi.org/10.3168/jds.S0022-0302(74)84993-3).
  13. Henderson C.R., Comparison of alternative sire evaluation methods, *Journal of Animal Science*, 1975, Vol. 41, No. 3, pp. 760–770, <https://doi.org/10.2527/jas1975.413760x>.
  14. Henderson C.R., Best linear unbiased prediction of breeding values not in the model for records, *Journal of Dairy Science*, 1977, Vol. 60, No 5, pp. 783–787, [https://doi.org/10.3168/jds.S0022-0302\(77\)83935-0](https://doi.org/10.3168/jds.S0022-0302(77)83935-0).
  15. Schaeffer L.R., Application of random regression models in animal breeding, *Livestock Production Science*, 2004, Vol. 86, No. 1–3, pp. 35–45, [https://doi.org/10.1016/S0301-6226\(03\)00151-9](https://doi.org/10.1016/S0301-6226(03)00151-9).
  16. Henderson C.R., Quaas R.L., Multiple trait evaluation using relatives' records, *Journal of animal science*, 1976, Vol. 43, No. 6, pp. 1188–1197, <https://doi.org/10.2527/jas1976.4361188x>.
  17. Quaas R.L., Additive genetic model with groups and relationships, *Journal of Dairy Science*, 1988, Vol. 71, No. 5, pp. 1338–1345, [https://doi.org/10.3168/jds.S0022-0302\(88\)79691-5](https://doi.org/10.3168/jds.S0022-0302(88)79691-5).
  18. Yu C.R., Zou K.H., Carlsson M.O., Weerahandi S., Generalized estimation of the BLUP in mixed-effects models: A comparison with ML and REML, *Communications in Statistics-Simulation and Computation*, 2015, Vol. 44, No. 3, pp. 694–704, <https://doi.org/10.1080/03610918.2013.790445>.
  19. Schaeffer L.R. CR Henderson: Contributions to predicting genetic merit, *Journal of dairy science*, 1991, Vol. 74, No. 11, pp. 4052–4066, [https://doi.org/10.3168/jds.S0022-0302\(91\)78601-3](https://doi.org/10.3168/jds.S0022-0302(91)78601-3).
  20. Calus M. P.L., De Haas Y., Veerkamp R.F., Combining cow and bull reference populations to increase accuracy of genomic prediction and genome-wide association studies, *Journal of Dairy Science*, 2013, Vol. 96, No. 10, pp. 6703–6715, <https://doi.org/10.3168/jds.2012-6013>.
  21. Gianola D., Foulley J. L., Fernando R.L., Prediction of breeding values when variances are not known, *Génétique sélection évolution*, 1986, Vol. 18, No. 4, pp. 485–498.
  22. Yang H., Su G., Impact of phenotypic information of previous generations and depth of pedigree on estimates of genetic parameters and breeding values, *Livestock Science*, 2016, Vol. 187, pp. 61–67, <https://doi.org/10.1016/j.livsci.2016.03.001>.
  23. Henderson C.R., Equivalent linear models to reduce computations, *Journal of Dairy Science*, 1985, Vol. 68, No. 9, pp. 2267–2277, [https://doi.org/10.3168/jds.S0022-0302\(85\)81099-7](https://doi.org/10.3168/jds.S0022-0302(85)81099-7).
  24. Quaas R.L., Pollak E.J., Mixed model methodology for farm and ranch beef cattle testing programs, *Journal of Animal Science*, 1980, Vol. 51, No. 6, pp. 1277–1287. <https://doi.org/10.2527/jas1981.5161277x>.
  25. Gianola D., Fernando R.L., Bayesian methods in animal breeding theory, *Journal of Animal Science*, 1986, Vol. 63, No. 1, pp. 217–244, <https://doi.org/10.2527/jas1986.631217x>.
  26. Henderson C.R., Use of all relatives in intraherd prediction of breeding values and producing abilities, *Journal of Dairy Science*, 1975, Vol. 58, No. 12, pp. 1910–1916, [https://doi.org/10.3168/jds.S0022-0302\(75\)84808-9](https://doi.org/10.3168/jds.S0022-0302(75)84808-9).
  27. Van Vleck L.D., Derivation of Henderson's method of incorporating artificial insemination sire evaluations into intraherd prediction of breeding values, *Journal of Dairy Science*, 1982, Vol. 65, No. 2, pp. 284–286, [https://doi.org/10.3168/jds.S0022-0302\(82\)82190-5](https://doi.org/10.3168/jds.S0022-0302(82)82190-5).
  28. Chen X., Liang H., Wang Y., Total positivity of recursive matrices, *Linear Algebra and its Applications*, 2015, Vol. 471, pp. 383–393. <https://doi.org/10.1016/j.laa.2015.01.009>.
  29. Misztal I., Legarra A., Invited review: efficient computation strategies in genomic selection, *Animal*, 2017, Vol. 11, No. 5, pp. 731–736, doi:10.1017/S1751731116002366.
  30. Misztal I., Aggrey S.E., Muir W.M., Experiences with a single-step genome evaluation, *Poultry science*, 2013, Vol. 92, No. 9, pp. 2530–2534, <https://doi.org/10.3382/ps.2012-02739>.
  31. Henderson C.R., Simple method to compute biases and mean squared errors of linear estimators and predictors in a selection model assuming multivariate normality, *Journal of Dairy Science*, 1988, Vol. 71, No. 11, pp. 3135–3142, [https://doi.org/10.3168/jds.S0022-0302\(88\)79914-2](https://doi.org/10.3168/jds.S0022-0302(88)79914-2).

32. Henderson C.R., A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values, *Biometrics*, 1976, pp. 69–83. <https://doi.org/10.2307/2529339>.
33. Darvasi A., Experimental strategies for the genetic dissection of complex traits in animal models, *Nature genetics*, 1998, Vol. 18, No. 1, pp. 19–24.
34. Pollak E.J., Quaas R.L., Definition of group effects in sire evaluation models, *Journal of Dairy Science*, 1983, Vol. 66, No. 7, pp. 1503–1509, [https://doi.org/10.3168/jds.S0022-0302\(83\)81965-1](https://doi.org/10.3168/jds.S0022-0302(83)81965-1).
35. Quaas R.L., Computing the diagonal elements and inverse of a large numerator relationship matrix, *Biometrics*, 1976, pp. 949–953, <https://doi.org/10.2307/2529279>.
36. Misztal I., Legarra A., Aguilar I., Using recursion to compute the inverse of the genomic relationship matrix, *Journal of dairy science*, 2014, Vol. 97, No. 6, pp. 3943–3952, <https://doi.org/10.3168/jds.2013-7752>.
37. Misztal I., Restricted maximum likelihood estimation of variance components in animal model using sparse matrix inversion and a supercomputer, *Journal of Dairy Science*, 1990, Vol. 73, No. 1, pp. 163–172, [https://doi.org/10.3168/jds.S0022-0302\(90\)78660-2](https://doi.org/10.3168/jds.S0022-0302(90)78660-2).
38. Henderson C.R., Best linear unbiased prediction using relationship matrices derived from selected base populations, *Journal of Dairy Science*, 1985, Vol. 68, No. 2, pp. 443–448, [https://doi.org/10.3168/jds.S0022-0302\(85\)80843-2](https://doi.org/10.3168/jds.S0022-0302(85)80843-2).
39. Eding H., Crooijmans R.P., Groenen M.A., Meuwissen T.H. Eding H. [et al.], Assessing the contribution of breeds to genetic diversity in conservation schemes, *Genetics Selection Evolution*, 2002, Vol. 34, No. 5, pp. 613–633, <https://doi.org/10.1051/gse:2002027>.
40. Lourenco D.A.L., Misztal I., Tsuruta S., Aguilar I., Ezra E., Ron M., Weller J.I., Lourenco D.A.L. [et al.], Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses, *Journal of dairy science*, 2014, Vol. 97, No. 3, pp. 1742–1752, <https://doi.org/10.3168/jds.2013-6916>.
41. Laodim T., Elzo M.A., Koonawootrittriron S., Suwanasopee T., Jattawa D., Genomic-polygenic and polygenic predictions for milk yield, fat yield, and age at first calving in Thai multibreed dairy population using genic and functional sets of genotypes, *Livestock Science*, 2019, Vol. 219, pp. 17–24, <https://doi.org/10.1016/j.livsci.2018.11.008>.
42. Zhou J., *Advances in Pedigree Analysis: Hardy-Weinberg Equilibrium, Strain Imputation, and Maternal Effects*, University of California, Los Angeles, 2011.
43. Ahlborn-Breier G., Hohenboken W.D., Additive and nonadditive genetic effects on milk production in dairy cattle: evidence for major individual, *Journal of Dairy science*, 1991, Vol. 74, No. 2, pp. 592–602, [https://doi.org/10.3168/jds.S0022-0302\(91\)78206-4](https://doi.org/10.3168/jds.S0022-0302(91)78206-4).
44. Smith S.P., Mäki-Tanila A., Genotypic covariance matrices and their inverses for models allowing dominance and inbreeding, *Genetics Selection Evolution*, 1990, Vol. 22, No. 1, pp. 65–91.
45. VanRaden P.M., Hoeschele I., Rapid inversion of additive by additive relationship matrices by including sire-dam combination effects, *Journal of dairy science*, 1991, Vol. 74, No. 2, pp. 570–579, [https://doi.org/10.3168/jds.S0022-0302\(91\)78204-0](https://doi.org/10.3168/jds.S0022-0302(91)78204-0).
46. Schaeffer L.R., CR Henderson: Contributions to predicting genetic merit, *Journal of dairy science*, 1991, Vol. 74, No. 11, pp. 4052–4066, [https://doi.org/10.3168/jds.S0022-0302\(91\)78601-3](https://doi.org/10.3168/jds.S0022-0302(91)78601-3).
47. Kahneman D., Thaler R.H., Anomalies: Utility maximization and experienced utility, *Journal of economic perspectives*, 2006, Vol. 20, No. 1, pp. 221–234, DOI: 10.1257/089533006776526076.
48. Koivula M., Strandén I., Pösö J., Aamand G.P., Mäntysaari E.A., Koivula M. [et al.], Single-step genomic evaluation using multitrait random regression model and test-day data, *Journal of Dairy Science*, 2015, Vol. 98, No. 4, pp. 2775–2784, <https://doi.org/10.3168/jds.2014-8975>.
49. Gianola D., Fernando R.L., Im S., Foulley J.L., Gianola D. [et al.], Likelihood estimation of quantitative genetic parameters when selection occurs: models and problems, *Genome*, 1989, Vol. 31, No. 2, pp. 768–777, <https://doi.org/10.1139/g89-136>.
50. Gianola D., Im S., Fernando R.L., Prediction of breeding value under Henderson’s selection model: a revisitation, *Journal of dairy science*, 1988, Vol. 71, No. 10, pp. 2790–2798., [https://doi.org/10.3168/jds.S0022-0302\(88\)79873-2](https://doi.org/10.3168/jds.S0022-0302(88)79873-2).

51. Legarra A., Christensen O.F., Aguilar I., Misztal I., Legarra A. [et al.], Single Step, a general approach for genomic selection, *Livestock Science*, 2014, Vol. 166, pp. 54–65, <https://doi.org/10.1016/j.livsci.2014.04.029>.