

Article

A Note on the Conditioning of the \mathbf{H}^{-1} Matrix Used in Single-Step GBLUP

Mohammad Ali Nilforooshan 

Livestock Improvement Corporation, Private Bag 3016, Hamilton 3240, New Zealand;
mohammad.nilforooshan@lic.co.nz

Simple Summary: Compared to BLUP, in single-step genomic BLUP, $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ is added to the inverse of the pedigree relationship matrix (\mathbf{A}^{-1}), forming \mathbf{H}^{-1} , where \mathbf{G} is the genomic relationship matrix, and \mathbf{A}_{22} is the block of \mathbf{A} for genotyped animals. Incompatibility between \mathbf{G} and \mathbf{A} may cause inflated genetic variance. Blending and tuning \mathbf{G} with \mathbf{A}_{22} partially solves the problem. However, conditioning \mathbf{H}^{-1} might still be needed, which is usually performed via $\tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$. This may violate the properties upon which \mathbf{H} is built. Alternative ways of weighting the \mathbf{H}^{-1} components are presented to prevent/minimise violations of the properties of \mathbf{H} .

Abstract: The single-step genomic BLUP (ssGBLUP) is used worldwide for the simultaneous genetic evaluation of genotyped and non-genotyped animals. It is easily extendible to all BLUP models by replacing the pedigree-based additive genetic relationship matrix (\mathbf{A}) with an augmented pedigree-genomic relationship matrix (\mathbf{H}). Theoretically, \mathbf{H} does not introduce any artificially inflated variance. However, inflated genetic variances have been observed due to the incomparability between the genomic relationship matrix (\mathbf{G}) and \mathbf{A} used in \mathbf{H} . Usually, \mathbf{G} is blended and tuned with \mathbf{A}_{22} (the block of \mathbf{A} for genotyped animals) to improve its numerical condition and compatibility. If deflation/inflation is still needed, a common approach is weighting $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ in the form of $\tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$, added to \mathbf{A}^{-1} to form \mathbf{H}^{-1} . In some situations, this can violate the conditional properties upon which \mathbf{H} is built. Different ways of weighting the \mathbf{H}^{-1} components (\mathbf{A}^{-1} , \mathbf{G}^{-1} , \mathbf{A}_{22}^{-1} , and \mathbf{H}^{-1} itself) were studied to avoid/minimise the violations of the conditional properties of \mathbf{H} . Data were simulated on ten populations and twenty generations. Responses to weighting different components of \mathbf{H}^{-1} were measured in terms of the regression of phenotypes on the estimated breeding values (the lower the slope, the higher the inflation) and the correlation between phenotypes and the estimated breeding values (predictive ability). Increasing the weight on \mathbf{H}^{-1} increased the inflation. The responses to weighting \mathbf{G}^{-1} were similar to those for \mathbf{H}^{-1} . Increasing the weight on \mathbf{A}^{-1} (together with \mathbf{A}_{22}^{-1}) was not influential and slightly increased the inflation. Predictive ability is a direct function of the slope of the regression line and followed similar trends. Responses to weighting $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ depend on the inflation/deflation of evaluations from \mathbf{A}^{-1} to \mathbf{H}^{-1} and the compatibility of the two matrices with the heritability used in the model. One possibility is a combination of weighting $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ and weighting \mathbf{H}^{-1} . Given recent advances in ssGBLUP, conditioning \mathbf{H}^{-1} might become an interim solution from the past and then not be needed in the future.

Keywords: conditional property; inflated; relationship matrix; single-step GBLUP; weighting



Citation: Nilforooshan, M.A. A Note on the Conditioning of the \mathbf{H}^{-1} Matrix Used in Single-Step GBLUP. *Animals* **2022**, *12*, 3208. <https://doi.org/10.3390/ani12223208>

Academic Editor: Michael E. Davis

Received: 6 October 2022

Accepted: 17 November 2022

Published: 19 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The unified genetic evaluation of genotyped and non-genotyped animals has been of great interest. In an initial attempt, Misztal et al. [1] suggested a unified pedigree (\mathbf{A}) and genomic (\mathbf{G}) relationship matrix (\mathbf{H}_{ini}), in which genomic relationships between genotyped animals replaced their pedigree relationship coefficients in \mathbf{A} . Denoting non-genotyped and genotyped animals with 1 and 2:

$$\mathbf{H}_{\text{ini}} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{G} \end{bmatrix} = \mathbf{A} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} - \mathbf{A}_{22} \end{bmatrix}. \quad (1)$$

This relationship matrix did not condition the distributions of breeding values for genotyped and non-genotyped animals on each other, leading to incoherencies in the joint distribution of genetic values for genotyped and non-genotyped animals. Legarra et al. [2] presented an augmented (\mathbf{A} and \mathbf{G}) relationship matrix in which the genetic values of non-genotyped animals were conditioned to the genetic values of genotyped animals. The resulting matrix was:

$$\begin{aligned} \mathbf{H} &= \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} \\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} \end{bmatrix} \\ &= \mathbf{A} + \begin{bmatrix} \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22}) \\ (\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} - \mathbf{A}_{22} \end{bmatrix}, \end{aligned} \quad (2)$$

which can be simplified to any of the following:

$$\mathbf{H} = \begin{bmatrix} \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{I} \end{bmatrix} \mathbf{G} \begin{bmatrix} \mathbf{A}_{22}^{-1}\mathbf{A}_{21} \\ \mathbf{I} \end{bmatrix}, \quad (3)$$

$$\mathbf{H} = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{I} \end{bmatrix} \mathbf{G} \begin{bmatrix} \mathbf{A}_{22}^{-1}\mathbf{A}_{21} \\ \mathbf{I} \end{bmatrix}, \quad (4)$$

$$\mathbf{H} = \mathbf{A} + \begin{bmatrix} \mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{I} \end{bmatrix} (\mathbf{G} - \mathbf{A}_{22}) \begin{bmatrix} \mathbf{A}_{22}^{-1}\mathbf{A}_{21} \\ \mathbf{I} \end{bmatrix}. \quad (5)$$

In matrix \mathbf{H} , the genomic information in \mathbf{G} influences the relationships between non-genotyped and genotyped animals and among non-genotyped animals. Later, it was discovered that \mathbf{H}^{-1} can be indirectly obtained without forming and inverting \mathbf{H} [3,4].

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{H}^{11} & \mathbf{H}^{12} \\ \mathbf{H}^{21} & \mathbf{H}^{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{H}^{22} \end{bmatrix} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}. \quad (6)$$

Note that:

$$\begin{aligned} \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} &= \mathbf{H}_{22}^{-1} - \mathbf{A}_{22}^{-1} \\ &= \mathbf{H}^{22} - \mathbf{H}^{21}(\mathbf{H}^{11})^{-1}\mathbf{H}^{12} - \mathbf{A}_{22}^{-1} + \mathbf{A}^{21}(\mathbf{A}^{11})^{-1}\mathbf{A}^{12} \\ &= \mathbf{H}^{22} - \mathbf{A}^{21}(\mathbf{A}^{11})^{-1}\mathbf{A}^{12} - \mathbf{A}_{22}^{-1} + \mathbf{A}^{21}(\mathbf{A}^{11})^{-1}\mathbf{A}^{12} \\ &= \mathbf{H}^{22} - \mathbf{A}_{22}^{-1}. \end{aligned}$$

Matrix \mathbf{G} is not always full-rank (e.g., when the number of genotyped animals is greater than the number of loci or when there are duplicated genotypes, such as for identical twins). To force \mathbf{G} to be positive-definite and avoid large diagonal values of \mathbf{G}^{-1} due to the bad numerical condition of \mathbf{G} , the first step of conditioning \mathbf{G} often involves blending it with \mathbf{A}_{22} , which is always positive-definite (except in the existence of identical twins or clones [5]) and of good numerical conditions (i.e., $\mathbf{G} \leftarrow (1 - k)\mathbf{G} + k\mathbf{A}_{22}$, $0 < k < 1$). Blending introduces residual polygenic effects (genetic effects not captured by genetic markers) to the evaluation model without explicitly modelling it, where the scalar k is the ratio of the polygenic to the total additive genetic variance [6].

It is theoretically true that no artificially inflated variance is introduced via the \mathbf{H} matrix [2]. However, inflated genetic variances have been observed due to incompatibilities between \mathbf{G} and \mathbf{A}_{22} [6–9]. Incompatible \mathbf{G} and \mathbf{A}_{22} lead to incorrectly weighted pedigree

and genomic information [7,8]. Besides different distributions of \mathbf{G} and \mathbf{A}_{22} elements, incomplete and incorrect pedigree information, and genotyping and imputation errors, incompatibilities between \mathbf{G} and \mathbf{A}_{22} can be due to the non-random selection of genotyped animals [10], and the different bases and scales of the two matrices [7]. Matrices \mathbf{A}_{22} and \mathbf{G} regress data to different means. Matrix \mathbf{A}_{22} regresses solutions towards pedigree founders, animals in the pedigree with unknown parents or genetic groups if considered in the pedigree. On the other hand, \mathbf{G} regresses solutions toward a founder population comprising genotyped animals [5,10] since the real allele frequencies in the founder population are unknown. The average genetic merit of genotyped animals can be different from founders, especially in the presence of selection. Different approaches (referred to as tuning) have been used for correcting the base difference between \mathbf{G} and \mathbf{A}_{22} [7,11] and rebasing and scaling \mathbf{G} to improve its consistency with \mathbf{A}_{22} [10]. Those approaches were tested by Nilforooshan [9] on New Zealand Romney sheep. Christensen [8] and Gao et al. [6] tuned \mathbf{G} by regressing its averages to the averages of \mathbf{A}_{22} (Equations (7) and (8), respectively).

$$\begin{cases} \mu(\text{diag}(\mathbf{G}))\beta + \alpha = \mu(\text{diag}(\mathbf{A}_{22})) \\ \mu(\mathbf{G})\beta + \alpha = \mu(\mathbf{A}_{22}) \end{cases} \quad (7)$$

$$\begin{cases} \mu(\text{diag}(\mathbf{G}))\beta + \alpha = \mu(\text{diag}(\mathbf{A}_{22})) \\ \mu(\text{offdiag}(\mathbf{G}))\beta + \alpha = \mu(\text{offdiag}(\mathbf{A}_{22})) \end{cases} \quad (8)$$

The α and β scalars obtained by solving either of the equations above are used for transforming \mathbf{G} into $\beta\mathbf{G} + \alpha\mathbf{1}\mathbf{1}'$. Another solution proposed to tackle the problem of inflated genomic evaluations (i.e., an increased variance of genomic predictions) as a result of incorrectly scaled genomic and pedigree information was scaling $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ in the form of $\tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$ [3,12,13]. Applying $\tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$ is equivalent to transforming \mathbf{G} into $[\tau\mathbf{G}^{-1} - (\omega - 1)\mathbf{A}_{22}^{-1}]^{-1}$ [3,9], which equals $\mathbf{G}[(1 - \omega)\mathbf{G} + \tau\mathbf{A}_{22}]^{-1}\mathbf{A}_{22}$. It is also equivalent to replacing $\mathbf{G} - \mathbf{A}_{22}$ with $[\tau\mathbf{G}^{-1} - (\omega - 1)\mathbf{A}_{22}^{-1}]^{-1}$ in Equation (2) [12].

Reducing τ and ω values toward 0 brings \mathbf{G} closer to \mathbf{A}_{22} by bringing \mathbf{H}^{22} closer to \mathbf{A}^{22} . However, it is not easily quantifiable how \mathbf{G} and \mathbf{A}_{22} are proportionally combined. With τ and ω deviating from each other and 1, there is a risk of distorting the conditional properties of \mathbf{H} , because the changes made in \mathbf{H}^{22} are not reflected in other blocks of \mathbf{H}^{-1} . Whereas $1 - k$ and k are the commonly used blending coefficients of \mathbf{G} and \mathbf{A}_{22} , τ and ω are the commonly used blending coefficients of \mathbf{H}^{-1} and \mathbf{A}^{-1} . i.e.,

$$\mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1} \end{bmatrix} = \omega\mathbf{H}^{-1} + (1 - \omega)\mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\tau - \omega)\mathbf{G}^{-1} \end{bmatrix}. \quad (9)$$

Considering the above equation, there is no legitimate reason for ω being out of the boundary of 0 and 1, and $\tau - \omega$ being out of the boundary of -1 and 1. Martini et al. [12] studied τ ranging from 0.1 to 2, and ω ranging from -1 to 1 by steps of 0.1, leading to 420 analyses. Dealing with two parameters increases the number of analyses and validation tests in a two-dimensional space. It is assuming that the k coefficient has already been chosen and does not need to be validated. The most coherent approach for finding k is by restricted maximum likelihood (REML), as proposed by Christensen and Lund [4], rather than using empirical values by screening and validation.

Weighting \mathbf{G}^{-1} and \mathbf{A}_{22}^{-1} as $\tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$ has been used until recently [12–17]. Several improvements have been made to ssGBLUP [18] and the use of $\tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$ is declining. For example, one of the factors leading to the need for an ω considerably less than 1 was that inbreeding coefficients were considered in \mathbf{A}_{22}^{-1} but not in \mathbf{A}^{-1} [19]. The aim of this study was to communicate the problems that might occur using $\tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$, and investigate the possible solutions for weighting the \mathbf{H}^{-1} components if the modifications in \mathbf{G} are not satisfactory and the weighting of the \mathbf{H}^{-1} components is still needed for the deflation/inflation of genomic breeding values.

2. Methods

2.1. Possible Problems with $\tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$

The $(\tau - \omega)\mathbf{G}^{-1}$ matrix in Equation (9) is unconditional and not reflected in the other blocks of \mathbf{H}^{-1} . As such, some combinations of $\tau \neq \omega$ potentially distort the conditional properties of \mathbf{H} . However, any $\tau = \omega$ ranging from 0 to 1 is legitimate and can be considered as a blending of \mathbf{H}^{-1} and \mathbf{A}^{-1} . While it might make sense to weight \mathbf{G}^{-1} and \mathbf{A}_{22}^{-1} to bring them closer to each other and make them more compatible, weighting \mathbf{A}_{22}^{-1} causes incompatibility between \mathbf{A}_{22}^{-1} and \mathbf{A}^{-1} . Matrix \mathbf{H}^{-1} can also be written as:

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{I} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \end{bmatrix} \mathbf{A}^{11} \begin{bmatrix} \mathbf{I} \\ -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{bmatrix} \quad (10)$$

$$= \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} - \mathbf{A}_{22}^{-1} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{bmatrix}. \quad (11)$$

Weighting the components of $\begin{bmatrix} \mathbf{I} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \end{bmatrix} \mathbf{A}^{11} \begin{bmatrix} \mathbf{I} \\ -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{bmatrix}$ in Equation (10), the aim is to preserve the existing quadratic form. This study aimed to introduce weighting on the \mathbf{H}^{-1} components that are unlikely to introduce distortions to the conditional properties of \mathbf{H} . Weighting \mathbf{H}^{-1} can be performed on any of the following components:

1. \mathbf{H}^{-1} itself
2. $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$
3. \mathbf{G}^{-1}
4. \mathbf{A}^{-1}
5. \mathbf{A}^{11}
6. \mathbf{A}^{22}
7. \mathbf{A}_{22}^{-1}

2.2. Weighting \mathbf{H}^{-1}

This scenario is helpful when the heritability estimate (h^2) does not match the data or \mathbf{H}^{-1} . Heritability may change over time and as a result of selection. An outdated h^2 may differ from the current h^2 of the trait in the population. Estimating variance components is a computationally expensive process. The h^2 estimate might have been from a population subset or via a matrix other than \mathbf{H}^{-1} (\mathbf{A}^{-1} or \mathbf{G}^{-1}). Different relationship matrices contain different information and may result in different genetic variances and h^2 estimates [20].

2.3. Weighting $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$

Aguilar et al. [3] suggested using equal τ and ω . Weighting $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ by α is equivalent to $\alpha\mathbf{H}^{-1} + (1 - \alpha)\mathbf{A}^{-1}$.

2.4. Weighting \mathbf{G}^{-1}

This scenario can be understood as scaling the h^2 corresponding to \mathbf{G}^{-1} to the h^2 corresponding to \mathbf{A}^{-1} . No violation is made to the conditional properties of \mathbf{H}^{-1} , and weighting \mathbf{G}^{-1} by α is equivalent to using \mathbf{G}/α in \mathbf{H} . Therefore, instead of \mathbf{G} , \mathbf{G}/α is propagated through the blocks of \mathbf{H} . A \mathbf{G}/α more compatible with \mathbf{A}_{22} would bring \mathbf{G} closer to and more compatible with \mathbf{A} .

2.5. Weighting \mathbf{A}^{-1}

This scenario can be understood as scaling the h^2 corresponding to \mathbf{A}^{-1} to the h^2 corresponding to \mathbf{G}^{-1} . In response to \mathbf{A}^{-1} weighted by α , $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ in Equation (6) should be changed to $\mathbf{G}^{-1} - \alpha\mathbf{A}_{22}^{-1}$, which is equivalent to multiplying $\begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} - \mathbf{A}_{22}^{-1} \end{bmatrix}$

in Equation (11) by α . With an h^2 estimate based on pedigree information, weighting \mathbf{G}^{-1} is preferred over weighting \mathbf{A}^{-1} .

2.6. Weighting \mathbf{A}^{11}

Considering Equation (10), weighting \mathbf{A}^{11} is equivalent to weighting all the components of \mathbf{H}^{-1} , except \mathbf{G}^{-1} , similar to that of the weighting \mathbf{A}^{-1} scenario.

2.7. Weighting \mathbf{A}^{22}

Considering Equation (11), weighting \mathbf{A}^{22} should coincide with weighting the other blocks of \mathbf{A}^{-1} to preserve its conditional properties, as well as weighting \mathbf{A}_{22}^{-1} , similar to that of the weighting \mathbf{A}^{-1} scenario.

2.8. Weighting \mathbf{A}_{22}^{-1}

Considering Equation (10), weighting \mathbf{A}_{22}^{-1} is equivalent to:

$$\begin{aligned} \mathbf{H}^{-1} &= \begin{bmatrix} \mathbf{I} \\ -\sqrt{\alpha}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \end{bmatrix} \mathbf{A}^{11} \begin{bmatrix} \mathbf{I} \\ -\sqrt{\alpha}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}^{11} & \sqrt{\alpha}\mathbf{A}^{12} \\ \sqrt{\alpha}\mathbf{A}^{21} & \alpha(\mathbf{A}^{22} - \mathbf{A}_{22}^{-1}) \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sqrt{\alpha}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sqrt{\alpha}\mathbf{I} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \alpha\mathbf{A}_{22}^{-1} \end{bmatrix}. \quad (12) \end{aligned}$$

However, this is not recommended as it imposes a different pedigree-based h^2 on the genotyped and non-genotyped animals in \mathbf{A}^{-1} . Furthermore, as α becomes smaller, the relationships between genotyped and non-genotyped animals are weakened.

2.9. The Experiments

Since the scenarios of weighting \mathbf{A}^{11} and \mathbf{A}^{22} are equivalent to weighting \mathbf{A}^{-1} , and weighting \mathbf{A}_{22}^{-1} is not recommended, the four scenarios of weighting \mathbf{H}^{-1} , $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$, \mathbf{G}^{-1} , and \mathbf{A}^{-1} were tested. These scenarios were tested with α ranging from 0.8 to 1.2 to know the responses of each \mathbf{H}^{-1} conversion to the deviation of α from 1. Because weighting $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ requires α to be between 0 and 1, it was studied with α ranging from 0.8 to 1. Predictive ability was calculated as Pearson's correlation between the phenotypes and the estimated breeding values. Phenotypes were regressed on the estimated breeding values, where a lower slope means inflation and a higher slope means deflation.

3. Materials

Data were simulated for a species in a 1:1 sex ratio, litter size of 2, and generation overlap of 1. The pedigree, phenotypes, and genotypes were simulated using the R package *pedSimulate* [21]. Initially, ten generations were simulated, starting with a base generation (F0) of 100 animals (50 of each sex). No non-random pre-mating mortality or selection was applied to F0. Genotypes were simulated on 5000 markers, and allele frequencies were sampled from a uniform distribution ranging from 0.1 to 0.9. Marker (allele substitution) effects were simulated from a gamma distribution with shape and rate parameters equal to 2. The distribution was rebased to have a mean of 0 and scaled to create a variance of (true) marker breeding values in F0, $\sigma_g^2 = 9$. Residual polygenic and environment (residual) effects were simulated from normal distributions with variances $\sigma_a^2 = 1$ and $\sigma_e^2 = 30$, respectively.

Following F0, half of the males were mated to half of the females, which were all randomly selected and mated. Where the numbers of mating animals per sex were not equal, the sex with the higher number of animals underwent random selection to match the number of animals of the opposite sex. These ten generations were followed by ten more generations, in which 50% of male candidates (to become sires of the next generation)

were selected for their marker breeding value and mated to the same number of randomly selected females. Genotypes in each subsequent generation were obtained by combining sampled gametes from the parents' genotypes.

Phenotypes were calculated as $\mathbf{y} = \mu\mathbf{1} + \mathbf{g} + \mathbf{a} + \mathbf{e}$, where μ is the population mean, and \mathbf{g} , \mathbf{a} , and \mathbf{e} are the vectors of effects corresponding to σ_g^2 , σ_a^2 , and σ_e^2 . Genotypes before F8 and phenotypes for the last generation (F19) and before F7 were set to missing. Randomly, 5% of the known dams and 5% of the known sires (after F0) were set to missing. As such, missing pedigree and phenotype information, genomic pre-selection, and base and scale deviations between \mathbf{A} and \mathbf{G} were accommodated in the simulation. Data simulation was repeated ten times to reduce the possibility of observing the results specific to a dataset.

No fixed effect was simulated, and the data were analysed using the following mixed model equations:

$$\begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{1} & \mathbf{Z}'\mathbf{Z} + \mathbf{H}^{-1} \frac{\sigma_e^2}{\sigma_g^2 + \sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \Sigma \mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}, \tag{13}$$

where \mathbf{Z} is the matrix relating phenotypes to animals, $\mathbf{1}$ and $\hat{\mathbf{u}}$ are the vectors of ones and predicted breeding values, and $\hat{\mu}$ is the mean estimate. Matrix \mathbf{G} was used in \mathbf{H}^{-1} and built according to method 1 of VanRaden [5], where $\mathbf{G} = \mathbf{W}\mathbf{W}' / 2 \sum p(1-p)$, \mathbf{W} is the centred and scaled genotype matrix, and p is the marker allele frequency. Markers with minor allele frequency below 0.02 were discarded before calculating \mathbf{G} . Then, \mathbf{G} was blended as $\mathbf{G} \leftarrow 0.9\mathbf{G} + 0.1\mathbf{A}_{22}$.

4. Results

The simulated pedigrees had a population size of 2162.8 ± 358.3 ($\mu \pm \text{sd}$), 1326.4 ± 298.2 genotypes, 1324.6 ± 277.2 phenotypes, 1074.7 ± 156.8 males, and 1088.1 ± 202.9 females. Inflation and predictive ability estimates over the ten simulated pedigrees were averaged and presented (Figures 1 and 2).

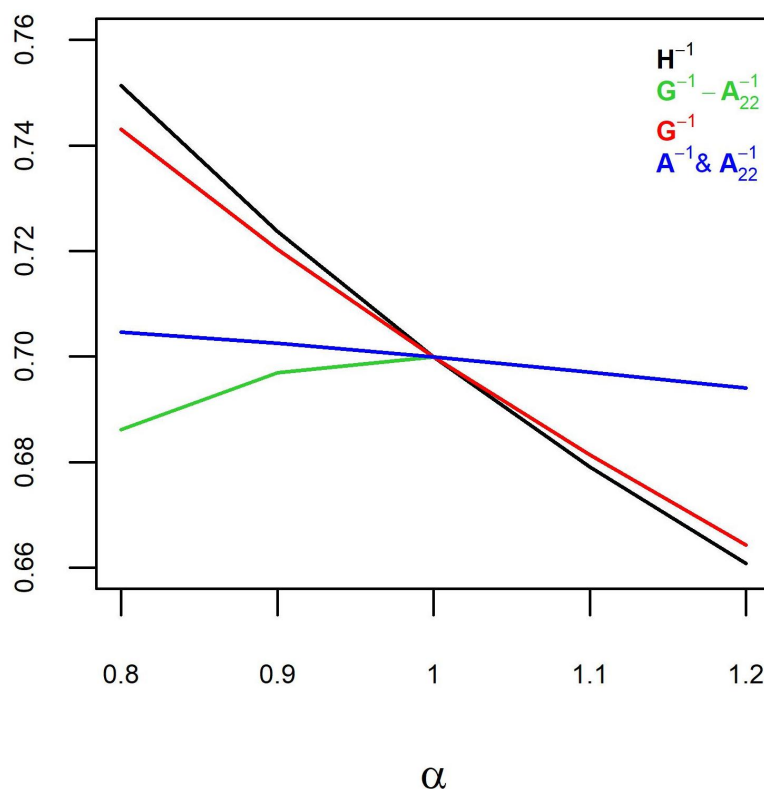


Figure 1. Regression coefficients of the phenotypes on genomic breeding values for different components of \mathbf{H}^{-1} weighted by α . Each data point is an average of ten observations for the simulated populations.

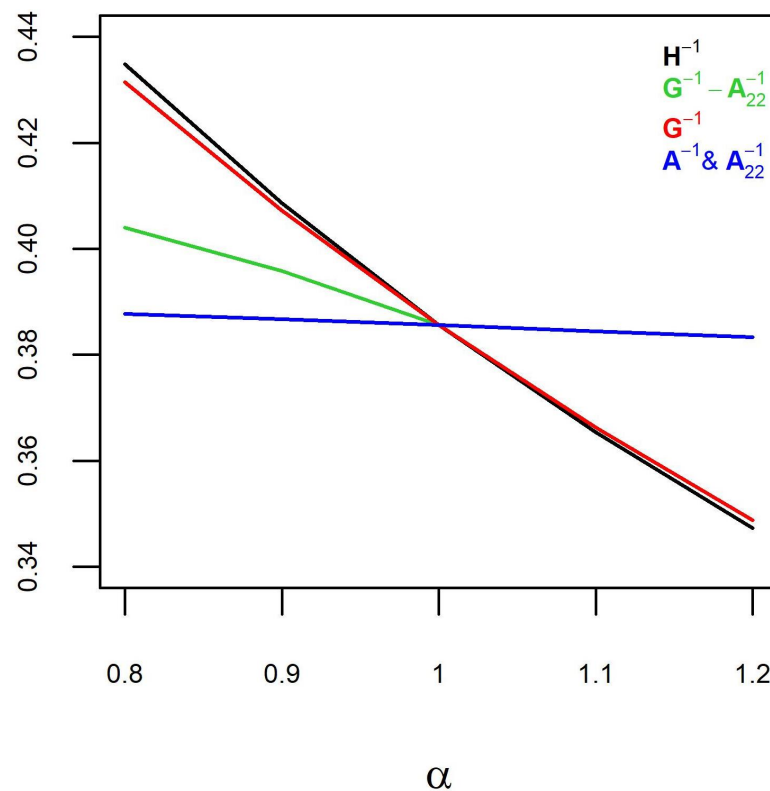


Figure 2. Correlation coefficients between phenotypes and genomic breeding values for different components of \mathbf{H}^{-1} weighted by α . Each data point is an average of ten observations for the simulated populations.

Different \mathbf{H}^{-1} components were weighted by α ranging from 0.8 to 1.2, except for $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$, where α ranged from 0.8 to 1. Weighting \mathbf{H}^{-1} and \mathbf{G}^{-1} showed similar trends for inflation (Figure 1) and predictive ability (Figure 2), with the slope of the trends being slightly less for \mathbf{G}^{-1} compared to \mathbf{H}^{-1} . Weighting \mathbf{A}^{-1} (accompanied by weighting \mathbf{A}_{22}^{-1}) showed slightly decreasing trends, with the regression slope decreasing by 0.01 (i.e., inflation increasing by 0.01) and the predictive ability decreasing by 4.4×10^{-3} over the range of α . The inflation and prediction ability increased by weighting $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ with α decreasing from 1 to 0.8.

5. Discussion

Matrices \mathbf{G} and \mathbf{A}_{22} indicate different means and variances for genotyped animals. This can cause differently scaled genomic and pedigree information in \mathbf{H}^{-1} [3]. Usually, \mathbf{G} is blended and tuned (rebased and scaled) with \mathbf{A}_{22} . If genomic breeding values are still inflated, a complementary weighting of $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ might be needed. A common practice is to weight using $\tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$. It was shown that some $\tau \neq \omega$ combinations are likely to distort the properties of \mathbf{H} that provide conditionality between the breeding values of genotyped and non-genotyped animals. Other ways of weighting the components of \mathbf{H}^{-1} were presented that are unlikely to distort the conditional properties of \mathbf{H} .

Weighting \mathbf{H}^{-1} with $\alpha > 1$ is equivalent to reducing h^2 and increasing inflation due to increased dispersion. It is equivalent to adding $(1 - \alpha)/\alpha$ to $1/h^2$ or weighting the genetic variance by $1/\alpha$. Due to selection, h^2 can be lower than expected. The h^2 reduction is expected to be greater due to genomic selection. Change of genetic variance by genomic selection is propagated from \mathbf{G} throughout \mathbf{H} . The predictive ability declined with increasing α (Figure 2), which might be concerning. However, predictive ability is a direct function of the slope of the regression line (Figure 1). Therefore, the slope of the regression line (inflation) should be the main concern.

Weighting \mathbf{A}^{-1} (accompanied by weighting \mathbf{A}_{22}^{-1}) did not influence inflation and predictive ability. Predictive ability and the slope of the regression line decreased slightly (inflation increased slightly) over the increase in α . The reason for this is likely that \mathbf{H} is a genomic relationship matrix extended from \mathbf{G} for genotyped animals to non-genotyped animals via the $\mathbf{A}_{12}\mathbf{A}_{22}^{-1}$ coefficients (Equations (2)–(5)). As such, \mathbf{G} is more influential in defining the variances in \mathbf{H} than \mathbf{A} . This was confirmed by similar trends for weighting \mathbf{G}^{-1} and \mathbf{H}^{-1} (Figures 1 and 2). The slopes of the regression line (inflation) and predictive ability were slightly steeper for \mathbf{H}^{-1} than for \mathbf{G}^{-1} , and that was a result of the combined weighting of \mathbf{G}^{-1} , \mathbf{A}^{-1} and \mathbf{A}_{22}^{-1} . Weighting $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ by $\alpha < 1$ increased the inflation but at a lower rate than weighting \mathbf{H}^{-1} or \mathbf{G}^{-1} with $\alpha > 1$.

The inflation results are expected to be valid for other data as weighting \mathbf{H}^{-1} or its components is equivalent to inversely weighting the genetic variance, regardless of the data. The exception is weighting $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$. Whether weighting $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ with a larger α results in inflation or deflation depends on whether using \mathbf{H}^{-1} instead of \mathbf{A}^{-1} results in inflation or deflation. If using \mathbf{H}^{-1} results in inflation, then weighting $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ with a larger α (more emphasis on \mathbf{H}^{-1} than \mathbf{A}^{-1}) results in greater inflation. The predictive ability improved by weighting $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ with α decreasing from 1 to 0.8. Generally, predictive ability increases by the increase in the slope of the regression line. Notice that the predictive ability ignoring inflation can be misleading. Since the trends for prediction ability and the slope of the regression line were in opposite directions for weighting $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$, it shows that the predictive ability benefited from blending \mathbf{H}^{-1} and \mathbf{A}^{-1} , mainly because the h^2 was more compatible with a blended \mathbf{H}^{-1} and \mathbf{A}^{-1} than with \mathbf{H}^{-1} .

This study does not completely rule out using $\tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$. However, weighting \mathbf{H}^{-1} components should meet specific conditions to avoid/minimise violating the conditional properties of \mathbf{H} . As such,

$$\begin{gathered} \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \alpha(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}) \end{bmatrix}, \\ \tau \left(\mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \alpha(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}) \end{bmatrix} \right), \\ \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \alpha\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}, \end{gathered}$$

and $\alpha\mathbf{H}^{-1}$ are better alternatives to $\tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$. By definition, none of these four options are better than the others. However, achieving good compatibility between the resulting \mathbf{H}^{-1} and h^2 without blending \mathbf{H}^{-1} and \mathbf{A}^{-1} at a high rate (low emphasis on genomic information) is important.

Concerning pedigree and genomic errors, regardless of the emphasis given to pedigree and genomic information, genotype errors propagate through non-genotyped animals, and pedigree errors incorrectly and insufficiently propagate genotype information through non-genotyped animals. Therefore, the correctness and the completeness of pedigree and genomic information are vital for accurate and unbiased ssGBLUP evaluations.

Future research may focus on changing genetic parameters over time or across populations in genomic predictions. It is possible to reduce inflation in genomic predictions for young animals by using smaller additive genetic variances. This can be done by replacing \mathbf{H}^{-1} with $\mathbf{D}\mathbf{H}^{-1}\mathbf{D}$. Considering no overall weight on \mathbf{H}^{-1} : $\sum \mathbf{D}\mathbf{H}^{-1}\mathbf{D} = \sum \mathbf{H}^{-1}$. Matrix \mathbf{D} is a diagonal matrix of positive values descending in function of the animal’s age. The researcher would need to decide the $\min(\mathbf{d}) \leq \frac{\sigma_e^2}{\sigma_s^2} \leq \max(\mathbf{d})$ range, where $\mathbf{d} = \text{diag}(\mathbf{D})$. With recent advances in ssGBLUP (mentioned by Misztal et al. [18]), which improve the compatibility between \mathbf{A} and \mathbf{G} , conditioning \mathbf{H}^{-1} might become an interim solution from the past or be reduced to only weighting \mathbf{H}^{-1} .

Funding: This work was supported by the NZ Ministry for Primary Industries, SFF Futures Programme: Resilient Dairy-Innovative breeding for a sustainable dairy future (grant number PGP06-17006).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data, code, and the results supporting the findings of this study are openly available in Mendeley Data at doi:10.17632/cn9jzpj7fg.1 [22].

Conflicts of Interest: M.A.N. is employed at Livestock Improvement Corporation, Hamilton, New Zealand. He declares that the research was conducted in the absence of any commercial or financial interest.

References

1. Misztal, I.; Legarra, A.; Aguilar, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* **2009**, *92*, 4648–4655. [CrossRef] [PubMed]
2. Legarra, A.; Aguilar, I.; Misztal, I. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* **2009**, *92*, 4656–4663. [CrossRef] [PubMed]
3. Aguilar, I.; Misztal, I.; Johnson, D.L.; Legarra, A.; Tsuruta, S.; Lawlor, T.J. Hot topic: A unified approach to utilise phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* **2010**, *93*, 743–752. [CrossRef] [PubMed]
4. Christensen, O.F.; Lund, M.S. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* **2010**, *42*, 2. [CrossRef] [PubMed]
5. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef] [PubMed]
6. Gao, H.; Christensen, O.F.; Madsen, P.; Nielsen, U.S.; Zhang, Y.; Lund, M.S.; Su, G. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genet. Sel. Evol.* **2012**, *44*, 8. [CrossRef] [PubMed]
7. Forni, S.; Aguilar, I.; Misztal, I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet. Sel. Evol.* **2011**, *43*, 1. [CrossRef] [PubMed]
8. Christensen, O.F. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. *Genet. Sel. Evol.* **2012**, *44*, 37. [CrossRef] [PubMed]
9. Nilforooshan, M.A. Application of single-step GBLUP in New Zealand Romney sheep. *Anim. Prod. Sci.* **2020**, *60*, 1139–1144. [CrossRef]
10. Vitezica, Z.G.; Aguilar, I.; Misztal, I.; Legarra, A. Bias in genomic predictions for populations under selection. *Genet. Res.* **2011**, *93*, 357–366. [CrossRef] [PubMed]
11. Chen, C.Y.; Misztal, I.; Aguilar, I.; Legarra, A.; Muir, W.M. Effect of different genomic relationship matrices on accuracy and scale. *J. Anim. Sci.* **2011**, *89*, 2673–2679. [CrossRef] [PubMed]
12. Martini, J.W.R.; Schrauf, M.F.; Garcia-Baccino, C.A.; Pimentel, E.C.G.; Munilla, S.; Rogberg-Muñoz, A.; Cantet, R.J.C.; Reimer, C.; Gao, N.; Wimmer, V.; et al. The effect of the H^{-1} scaling factors τ and ω on the structure of H in the single-step procedure. *Genet. Sel. Evol.* **2018**, *50*, 16. [CrossRef] [PubMed]
13. Misztal, I.; Aguilar, I.; Legarra, A.; Lawlor, T.J. Choice of parameters for single-step genomic evaluation for type. In Proceedings of the 61st Annual EAAP Meeting, Heraklion, Greece, 23–27 August 2010; p. 357.
14. Kang, H.; Ning, C.; Zhou, L.; Zhang, S.; Yan, Q.; Liu, J.-F. Short communication: Single-step genomic evaluation of milk production traits using multiple-trait random regression model in Chinese Holsteins. *J. Dairy Sci.* **2018**, *101*, 11143–11149. [CrossRef] [PubMed]
15. Imai, A.; Kuniga, T.; Yoshioka, T.; Nonaka, K.; Mitani, N.; Fukamachi, H.; Hiehata, N.; Yamamoto, M.; Hayashiet, T. Single-step genomic prediction of fruit-quality traits using phenotypic records of non-genotyped relatives in citrus. *PLoS ONE* **2019**, *14*, e0221880. [CrossRef] [PubMed]
16. Alvarenga, A.B.; Veroneze, R.; Oliveira, H.R.; Marques, D.B.D.; Lopes, P.S.; Silva, F.F.; Brito, L.F. Comparing alternative single-step GBLUP approaches and training population designs for genomic evaluation of crossbred animals. *Front. Genet.* **2020**, *11*, 263. [CrossRef] [PubMed]
17. Fu, C.; Ostensen, T.; Christensen, O.F.; Xiang, T. Single-step genomic evaluation with metafounders for feed conversion ratio and average daily gain in Danish Landrace and Yorkshire pigs. *Genet. Sel. Evol.* **2021**, *53*, 79. [CrossRef] [PubMed]
18. Misztal, I.; Lourenco, D.; Tsuruta, S.; Aguilar, I.; Masuda, Y.; Bermann, M.; Cesarani, A.; Legarra, A. How ssGBLUP became suitable for national dairy cattle evaluations. In Proceedings of the 12th World Congress on Genetics Applied to Livestock Production, Rotterdam, The Netherlands, 3–8 July 2022; p. 357. Available online: https://www.wageningenacademic.com/pb-assets/wagen/WCGALP2022/52_009.pdf (accessed on 5 October 2022).
19. Lourenco, D.A.L.; Legarra, A.; Tsuruta, S.; Masuda, Y.; Aguilar, I.; Misztal, I. Single-step genomic evaluations from theory to practice: Using SNP chips and sequence data in BLUPF90. *Genes* **2020**, *11*, 790. [CrossRef] [PubMed]

20. Legarra, A. Comparing estimates of genetic variance across different relationship models. *Theor. Pop. Biol.* **2016**, *107*, 26–60. [10.1016/j.tpb.2015.08.005](https://doi.org/10.1016/j.tpb.2015.08.005) [CrossRef] [PubMed]
21. Nilforooshan, M.A. pedSimulate—An R package for simulating pedigree, genetic merit, phenotype, and genotype data. *R. Bras. Zootec.* **2022**, *51*, e20210131. [CrossRef]
22. Nilforooshan, M.A. Code & Data—A Note on the Conditioning of the \mathbf{H}^{-1} Matrix Used in Single-Step GBLUP. Mendeley Data V1. 2022. Available online: <https://doi.org/10.17632/cn9jzpj7fg.1> (accessed on 15 November 2022).