Optimum and Decorrelated Constrained Multistage Linear Phenotypic Selection Indices Theory

J. Jesus Cerón-Rojas, Fernando H. Toledo, and Jose Crossa★

ABSTRACT

Some authors have evaluated the unconstrained *optimum* and *decorrelated* multistage linear phenotypic selection indices (OMLPSI and DMLPSI, respectively) theory. We extended this index theory to the constrained multistage linear phenotypic selection index context, where we denoted OMLPSI and DMLPSI as OCMLPSI and DCMLPSI, respectively. The OCMLPSI (DCMLPSI) is the most general multistage index and includes the OMLPSI (DMLPSI) as a particular case. The OCMLPSI (DCMLPSI) predicts the individual net genetic merit at different individual ages and allows imposing constraints on the genetic gains to make some traits change their mean values based on a predetermined level, while the rest of them remain without restrictions. The OCMLPSI takes into consideration the index correlation values among stages, whereas the DCMLPSI imposes the restriction that the index correlation values among stages be null. The criteria to evaluate OCMLPSI efficiency vs. DCMLPSI efficiency were that the total response of each index must be lower than or equal to the single-stage constrained linear phenotypic selection index response and that the expected genetic gain per trait values should be similar to the constraints imposed by the breeder. We used one real and one simulated dataset to validate the efficiency of the indices. The results indicated that OCMLPSI accuracy when predicting the selection response and expected genetic gain per trait was higher than DCMLPSI accuracy when predicting them. Thus, breeders should use the OCMLPSI when making a phenotypic selection.

Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, 06600, Mexico City, Mexico. Received 13 Apr. 2019. Accepted 1 July 2019. ★Corresponding author (j.crossa@cgiar.org). Assigned to Associate Editor Michael Casler.

Abbreviations: BW, body weight; DCMLPSI, decorrelated constrained multistage linear phenotypic selection index; DMLPSI, unconstrained decorrelated multistage linear phenotypic selection index; EW, egg weight; LPSI, single-stage linear phenotypic selection index; OCMLPSI, optimum constrained multistage linear phenotypic selection index; OMLPSI, unconstrained optimum multistage linear phenotypic selection index; NMV, multivariate normal distribution; QTL, quantitative trait locus; REML, restricted maximum likelihood; RL, rate of lay; SCLPSI, single-stage constrained linear phenotypic selection index; SM, age at sexual maturity.

In a two-stage context, Cerón-Rojas et al. (2019) described and evaluated the unconstrained *optimum* and *decorrelated* multistage linear phenotypic selection indices (OMLPSI and DMLPSI, respectively) theory and concluded that OMLPSI efficiency when predicting the net genetic merit was higher than the DMLPSI efficiency and that breeders should use the OMLPSI when making phenotypic selection. The main difference between the two indices is that although the OMLPSI takes into consideration the correlation values among stages when predicting the net genetic merit, the DMLPSI imposes the restriction that the correlation values among stages be null when it makes the prediction. The main characteristic of the OMLPSI (DMLPSI) in a two-stage context is that at Stage 1, OMLPSI (DMLPSI) is a partial index, but at Stage 2, it is a complete index. This selection procedure is called the *part and whole index selection method* (Young, 1964; Saxton, 1983) and is valid for any number of stages. The OMLPSI (DMLPSI) is more efficient than the independent culling method because it uses all available information at each

stage and incorporates the genetic correlations between traits in the prediction.

The OMLPSI (DMLPSI) combines the single-stage linear phenotypic selection index (LPSI) theory (Smith, 1936; Hazel, 1943) with the independent culling selection method (Cochran, 1951; Young, 1964; Cunningham, 1975; Xu and Muir, 1992) and is useful for selecting more than one trait in the multistage selection context. Breeders apply the OMLPSI (DMLPSI) mainly in animal and tree breeding where, due to early culling, OMLPSI (DMLPSI) is a cost-saving strategy for improving several traits because they do not need to measure all traits at each stage. The OMLPSI (DMLPSI) increases selection intensity on traits measured at an earlier age, and, with fixed facilities, this index selects a greater number of individuals at an earlier age (Xu and Muir, 1991, 1992).

The OMLPSI values may have a non-normal distribution after the first selection stage, and to derive selection intensities for more than two stages, this index requires numeric multiple integration techniques. To solve this problem, the DMLPSI minimizes the mean squared difference between the index and the net genetic merit at each stage under the restriction that the covariance between the DMLPSI values at different stages be zero, thus preventing the correlation between DMLPSI values at different stages. Under this restriction, truncation points and selection intensities can be determined for a fixed total proportion before the breeder carries out selection, and the selected individual index values after the first selection stage may be normally distributed (Xu and Muir, 1992). Nevertheless, due to the indicated restriction, the DMLPSI selection response and accuracy after the first stage could be lower than the OMLPSI selection response.

One additional problem with the OMLPSI (DMLPSI) expected genetic gain per trait (or multitrait selection response) is that its values can increase or decrease in a positive or negative direction without control. In the single-stage context, Kempthorne and Nordskog (1959) developed the restricted LPSI that allows imposing restrictions equal to zero on the expected genetic gain of some traits. Other authors (Mallard, 1972; Harville, 1975; Tallis, 1985) extended the Kempthorne and Nordskog (1959) approach and developed a single-stage constrained LPSI (SCLPSI) that attempts to make some traits change their expected genetic gain values based on a predetermined level while the rest of the traits remain without restrictions. Itoh and Yamada (1987) showed that in reality there is only one optimum SCLPSI; that is, the Mallard (1972), Harville (1975), and Tallis (1985) indices are the same. Xie and Xu (1997) and Cerón-Rojas and Crossa (2018, Chapter 9) extended the DMLPSI and OMLPSI to the constrained context, respectively. The Xie and Xu (1997) index, however, is not an optimum constrained multistage index because their approach is based on the single-stage Tallis (1962) constrained index theory, which is not an optimum index (see Cerón-Rojas and Crossa, 2018, Chapter 3, for details).

Based on the Mallard (1972) constrained phenotypic single-stage index theory, which is an optimum single-stage constrained index (see Cerón-Rojas and Crossa, 2018, Chapter 3, for details), in this work, we extend the OMLPSI and DMLPSI to the constrained multistage selection context. We will denote the OMLPSI and DMLPSI as OCMLPSI (optimum constrained multistage LPSI) and DCMLPSI (decorrelated constrained multistage LPSI), respectively. The main difference between the OCMLPSI and the DCMLPSI is that the OCMLPSI imposes only one restriction when solving the OMLPSI equations to obtain its vector of coefficients, whereas the DCMLPSI imposes two restrictions. The OCMLPSI solves the OMLPSI equations subject to the restriction that the covariance between the OCMLPSI and some linear combinations of the genotypes involved be equal to a vector of predetermined proportional gains (or constraints) imposed by the breeder, whereas the DCMLPSI imposes the additional restriction that the covariance between DCMLPSI values at different stages be zero. This additional restriction negatively affects the DCMLPSI selection response and expected genetic gain values per trait after the first stage.

One of the purposes of conducting a multistage selection is to reduce the cost and still obtain a reasonable gain. This means that the OCMLPSI and DCMLPSI could also be optimized with respect to aggregated economic gain and cost associated with obtain measures on each trait, but in this work, that problem was not considered. In a two-stage context, Namkoong (1970) has detailed how this last problem could be solved for the OMLPSI, whereas Xu and Muir (1992) have described that problem in the DMLPSI context.

We compared the relative efficiency of OCMLPSI and DCMLPSI under the assumption that the net genetic merit and the OCMLPSI and DCMLPSI values have joint multivariate normal distribution. We corroborated the normality assumption at Stage 2 using graphical methods and normality tests (Shapiro and Wilk, 1965; Mardia, 1980). Under this assumption, the regression of the net genetic merit on any linear function of the phenotypic values is linear (Kempthorne and Nordskog, 1959) and the selection response and expected genetic gain per trait results for two or more stages can be summarized arithmetically (Cochran, 1951; Young, 1964). We used two criteria to compare the efficiency of both indices. The first criterion was that the total selection response of each index must be lower than or equal to the SCLPSI selection response (Young, 1964; Saxton, 1983; Cerón-Rojas et al., 2019). The second criterion was that the expected genetic gain per trait values should be similar to the predetermined gains or constraints imposed by the breeder. We used one real

and one simulated dataset, each with four traits, to validate OCMLPSI efficiency vs. DCMLPSI efficiency. The results of both datasets indicated that the OCMLPSI is the most efficient index for predicting the net genetic merit, and its accuracy when predicting the selection response and estimating the expected genetic gain per trait was higher than the DCMLPSI accuracy when predicting the selection response and estimating the expected genetic gain per trait. Thus, breeders should use OCMLPSI when making a constrained phenotypic selection.

Results of this study are the first ones comparing (with real and simulated data) the relative efficiencies of the OCMLPSI vs. DCMLPSI using the total selection response and expected genetic gain pert trait as the main criteria to compare the efficiency of both indices.

MATERIALS AND METHODS
Methods
Objectives of the Constrained Multistage Linear Phenotypic Selection Indices

Let m_j be the population mean of the jth trait before selection. One of the main OCMLPSI (DCMLPSI) objectives is to change m_j to $m_j + d_j$, where d_j is the jth ($j = 1, 2, …, r$; r = the number of constrained traits) constrained trait or the jth predetermined proportional gain imposed by the breeder on the OCMLPSI (DCMLPSI) expected genetic gain per trait (Mallard, 1972; Cerón-Rojas and Crossa, 2018, Chapter 3). Additional OCMLPSI (DCMLPSI) objectives are (i) to maximize the selection response; (ii) to predict the net genetic merit ($H = \mathbf{w}'\mathbf{g}$, where $\mathbf{w}' = [w_1 \quad w_2 \quad … \quad w_n]$ and $\mathbf{g}' = [g_1 \quad g_2 \quad … \quad g_n]$ are $1 \times n$ vectors of economic weights and true unobservable breeding values, respectively); and (iii) to select individuals with the highest H values as parents of the next generation.

The Part and Whole Phenotypic Index Selection Method

Let $\mathbf{y}' = [y_1 \quad y_2 \quad … \quad y_n]$ be a $1 \times n$ vector of scores for n traits and assume that we can select only n_i of them at Stage i ($i = 1, 2, …, N$; N = number of stages) such that after N stages, $n = n_1 + n_2 + … + n_N$, where $n_i < N < n$. We can partition $\mathbf{y}$ into N subvectors as $\mathbf{y}' = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad … \quad \mathbf{x}_N]$, where $\mathbf{x}'_i = [y_1 \quad y_2 \quad … \quad y_{n_i}]$ is the subvector of $\mathbf{y}$ at Stage i ($i = 1, 2, …, N$). This means that at this stage, the ith index is $I_i = \beta_{i1}y_{i1} + \beta_{i2}y_{i2} + … + \beta_{in_i}y_{in_i} = \boldsymbol{\beta}'_i\mathbf{x}_i$, where $\boldsymbol{\beta}'_i = [\beta_{i1} \quad \beta_{i2} \quad … \quad \beta_{in_i}]$ is the index vector of coefficients, whereas $\mathbf{x}_i$ was defined earlier. Let

$$\mathbf{B}'_O = \begin{bmatrix} \boldsymbol{\beta}'_1 & 0 & … & 0 \\ \boldsymbol{\beta}'_1 & \boldsymbol{\beta}'_2 & … & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\beta}'_1 & \boldsymbol{\beta}'_2 & … & \boldsymbol{\beta}'_N \end{bmatrix}$$

be a transforming matrix; then, for each stage, we can construct an index as

$$\begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}'_1 & 0 & … & 0 \\ \boldsymbol{\beta}'_1 & \boldsymbol{\beta}'_2 & … & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\beta}'_1 & \boldsymbol{\beta}'_2 & … & \boldsymbol{\beta}'_N \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{bmatrix}$$

(Xu and Muir, 1992; Cerón-Rojas et al., 2019). This last result indicates that until Stage $N − 1$, each index is partial, but at Stage N, $I_N = \boldsymbol{\beta}'_1\mathbf{x}_1 + \boldsymbol{\beta}'_2\mathbf{x}_2 + … + \boldsymbol{\beta}'_N\mathbf{x}_N$ is a whole index.

Young (1964) called the foregoing procedure the *part and whole index selection method*. Xu and Muir (1992) called that selection procedure *selection index updating* because as traits become available, each subsequent index contains all traits available up to that stage. This method is more efficient than the independent culling selection method because it uses the genetic correlation among traits and all available information at each stage to predict the net genetic merit (Saxton, 1983). In addition, the independent culling selection method cannot impose constraints on the expected genetic gain of each trait, as the constrained index does.

Genotypic and Phenotypic Covariance Matrices

Let $\mathbf{g}' = [g_1 \quad g_2 \quad … \quad g_n]$, $\mathbf{x}'_i = [z_1 \quad z_2 \quad … \quad z_{n_i}]$, and $\mathbf{y}' = [y_1 \quad y_2 \quad … \quad y_n]$ be vectors, as defined in the above subsections. Thus, the genotypic covariance matrix of vectors $\mathbf{x}_i$ and $\mathbf{g}$ for N stages is

$$\begin{bmatrix} \text{Cov}(\mathbf{x}_1, \mathbf{g}) \\ \text{Cov}(\mathbf{x}_2, \mathbf{g}) \\ \vdots \\ \text{Cov}(\mathbf{x}_N, \mathbf{g}) \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \vdots \\ \mathbf{G}_N \end{bmatrix} = \mathbf{G}$$

whereas the phenotypic covariance matrix of vector $\mathbf{y}$ is $\text{Var}(\mathbf{y}) = \{\mathbf{P}_{ij}\} = \mathbf{P}$, where $\text{Cov}(\mathbf{x}_i, \mathbf{g}) = \mathbf{G}_i$ is the ith submatrix of $\mathbf{G}$, and $\mathbf{P}_{ij} = \text{Cov}(\mathbf{x}_i, \mathbf{x}_j)$ is the ijth ($i, j = 1, 2, …, N$) submatrix of $\mathbf{P}$.

To obtain the OCMLPSI (DCMLPSI) parameters, we need the following matrices:

$$\mathbf{Q}_{ii} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & … & \mathbf{P}_{1i} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & … & \mathbf{P}_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{i1} & \mathbf{P}_{i2} & … & \mathbf{P}_{ii} \end{bmatrix}, \quad \mathbf{A}_i = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \vdots \\ \mathbf{G}_i \end{bmatrix} \quad [1a]$$

which are submatrices of $\mathbf{P}$ and $\mathbf{G}$, respectively. In Appendix A (Eq. [A1] to [A3]), we describe a method to estimate $\mathbf{P}$ and $\mathbf{G}$.

Now suppose that the number of traits selected up to Stage $i − 1$ is $n_{i−1}$ and that at Stage i we select n_i traits, such that $n_i \leq n_{i−1}$ (or $n_{i−1} < n_i$). Then, according to the *part and whole index selection method*, at Stage i, we shall have $n_{i−1} + n_i$ traits. This means that the phenotypic covariance matrix $[\mathbf{Q}_{(i−1)i}]$ obtained with the $n_{i−1}$ traits selected at Stage $i − 1$ and the total $n_{i−1} + n_i$ traits will be of size $n_{i−1}(n_{i−1} + n_i)$ and can be written as

$$\mathbf{Q}_{(i−1)i} = \{s_{jc}\} \quad [1b]$$

where s_{jc} is the jcth phenotypic covariance value for $j = 1, 2, …, n_{i−1}$ and $c = 1, 2, …, (n_{i−1} + n_i)$. In addition, $n_{i−1}$ and ($n_{i−1}$

$+ n_i$) are the numbers of rows and columns of matrix $\mathbf{Q}_{(i-1)i}$, respectively. Equation [1b] indicates that $\mathbf{Q}_{(i-1)i}$ is a nonsquare and nonsymmetric matrix. Matrix $\mathbf{Q}_{(i-1)i}$ is useful for imposing the restrictions that make the DCMLPSI values independent among stages (see Cerón-Rojas et al., 2019, for details).

Selection Response at Stage i

At Stage i, the selection response (R_i) is the ith net genetic merit ($H = \mathbf{w}'\mathbf{g}$) mean of the selected population and can be written as

$$R_i = k_i \sigma_H \rho_{HI_i} \qquad [2]$$

where k_i is the selection intensity (Xu and Muir, 1992; Cerón-Rojas et al., 2019), $\sigma_H = \sqrt{\mathbf{w}'\mathbf{C}\mathbf{w}}$ is the standard deviation of $H = \mathbf{w}'\mathbf{g}$, Var($\mathbf{g}$) = $\mathbf{C}$ is the covariance matrix of $\mathbf{g}$, and ρ_{HI_i} is the correlation between $H = \mathbf{w}'\mathbf{g}$ and the index at Stage i ($I_i = \boldsymbol{\beta}'_i\mathbf{x}_i$). For N stages, the total selection response is $R_t = R_1 + R_2 + \dots + R_N$ (Cochran, 1951; Young, 1964). Equation [2] indicates that the genetic gain that can be achieved in R_i by selecting for several traits simultaneously within a population of animals or plants is the product of k_i, σ_H, and ρ_{HI_i} (Kempthorne and Nordskog, 1959). Selection intensity is limited by the rate of reproduction of each species, whereas σ_H is beyond human control; hence, the greatest opportunity for increasing selection progress is by ensuring that ρ_{HI_i} is as large as possible (Hazel 1943). Equation [2] is a useful criterion for comparing the efficiency of different types of indices to predict the net genetic merit ($H = \mathbf{w}'\mathbf{g}$; e.g., OCMLPSI efficiency vs. DCMLPSI efficiency). We would expect that the greater Eq. [2] is, the more effective OCMLPSI (DCMLPSI) is at predicting $H = \mathbf{w}'\mathbf{g}$. In the multistage selection index context, however, one main restriction is that the whole OCMLPSI (DCMLPSI) selection response be lower than or equal to the SCLPSI response (Saxton, 1983; Cerón-Rojas et al., 2019).

Expected Genetic Gain per Trait at Stage i

The expected genetic gain per trait at Stage i ($\mathbf{E}_i$, or multitrait selection response) is the covariance between the true breeding value vector ($\mathbf{g}$) and the $I_i = \boldsymbol{\beta}'_i\mathbf{x}_i$ value weighted by its standard deviation ($\sigma_{I_i} = \sqrt{\boldsymbol{\beta}'_i\mathbf{Q}_{ii}\boldsymbol{\beta}_i}$) and multiplied by the selection intensity (k_i), so that

$$\mathbf{E}_i = k_i \frac{\mathbf{A}'_i\boldsymbol{\beta}_i}{\sigma_{I_i}} \qquad [3]$$

We defined all the parameters of Eq. [3] previously. In the univariate and single-stage breeding scheme, Eq. [3] is the same as the selection response. For N stages, the total expected genetic gain per trait is $\mathbf{E}_t = \mathbf{E}_1 + \mathbf{E}_2 + \dots + \mathbf{E}_N$ (Cochran, 1951; Young, 1964).

In the OCMLPSI context, we will minimize the mean squared difference between the net genetic merit $H = \mathbf{w}'\mathbf{g}$ and the index $I_i = \boldsymbol{\beta}'_i\mathbf{x}_i$ {i.e., $E(H - I_i)^2$]} with respect to the vector of coefficients $\boldsymbol{\beta}_i$ ($i = 1, 2, \dots, N$) under the assumption that Eq. [3] values are equal to the d_j ($j = 1, 2, \dots, r$; r = number of constraints) values imposed by the breeder. The resulting vector of coefficients ($\boldsymbol{\beta}_i$) should maximize Eq. [2] and make the Eq. [3] values be near the d_j value. In the DCMLPSI context, however, it is necessary to impose the additional restriction that

the DCMLPSI values among stages are independent, as we shall see in the next two subsections.

The OCMLPSI Vector of Coefficients at Stage i

Let $\mathbf{d}' = [d_1\ d_2\ \dots\ d_r]$ be a vector $1 \times r$ of constraints or predetermined proportional gains per trait imposed by the breeder, and

$$\mathbf{D}' = \begin{bmatrix} d_r & 0 & \dots & 0 & -d_1 \\ 0 & d_r & \dots & 0 & -d_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & d_r & -d_{r-1} \end{bmatrix}$$

be a Mallard (1972) matrix of size $(r-1)r$, where d_j ($j = 1, 2, \dots, r$) is the jth element of vector $\mathbf{d}'$. In addition, let $\mathbf{U}'$ be a Kempthorne and Nordskog (1959) matrix $(n-r)n$ (n = number of traits and r = number of constraints) of 1s and 0s, where 1 indicates that the trait is constrained and 0 indicates that the trait has no constraints (see Cerón-Rojas and Crossa, 2018, Chapter 3, for details). According to the single-stage Mallard (1972) constrained index theory, to obtain the OCMLPSI vector of coefficients at Stage i, we need to minimize the mean squared difference between the net genetic merit $H = \mathbf{w}'\mathbf{g}$ and the index $I_i = \boldsymbol{\beta}'_i\mathbf{x}_i$ {i.e., i.e., $E(H - I_i)^2$]} under the restrictions $\mathbf{M}'_i\boldsymbol{\beta}_I = 0$, where $\mathbf{M}'_i = \mathbf{D}'\mathbf{U}'\mathbf{A}'_i$ and $\mathbf{A}'_i$ was defined in Eq. [1a].

Suppose that matrices $\mathbf{Q}_{ii}$, $\mathbf{U}$, and $\mathbf{A}'_i$, and vectors $\mathbf{d}$ and $\mathbf{w}$ are known at Stage i; then, it is necessary to minimize the function

$$f_O(\boldsymbol{\beta}_i, \mathbf{u}) = \boldsymbol{\beta}'_i\mathbf{Q}_{ii}\boldsymbol{\beta}_i - 2\mathbf{w}\mathbf{A}'_i\boldsymbol{\beta}_i + 2\mathbf{u}'\mathbf{M}'_i\boldsymbol{\beta}_i, \qquad [4]$$

with respect to the vector of coefficients $\boldsymbol{\beta}_i$ and the vector of Lagrange multipliers $\mathbf{u}' = [u_1\ u_2\ \dots\ u_{r-1}]$. The OCMLPSI vector of coefficients at Stage i is

$$\boldsymbol{\beta}_i = \mathbf{K}_{Oi}\boldsymbol{\delta}_i \qquad [5]$$

where $\boldsymbol{\delta}_i = \mathbf{Q}_{ii}^{-1}\mathbf{A}_i\mathbf{w}$, $\mathbf{Q}_{ii}^{-1}$ is the inverse of matrix $\mathbf{Q}_{ii}$, and $\mathbf{A}_i$ and $\mathbf{w}$ were defined earlier. In addition, $\mathbf{K}_{Oi} = [\mathbf{I}_i - \mathbf{F}_{Oi}]$, $\mathbf{F}_{Oi} = \mathbf{Q}_{ii}^{-1}\mathbf{M}_i(\mathbf{M}'_i\mathbf{Q}_{ii}^{-1}\mathbf{M}_i)^{-1}\mathbf{M}'_i$, and $\mathbf{I}_i$ is an identity matrix of the same size as $\mathbf{Q}_{ii}$. When $\mathbf{D} = \mathbf{U}$, the vector of coefficients of Eq. [5] imposes null restrictions, and when $\mathbf{D} = \mathbf{U}$ and $\mathbf{U}$ is a null matrix, Eq. [5] is equal to $\boldsymbol{\delta}_i = \mathbf{Q}_{ii}^{-1}\mathbf{A}_i\mathbf{w}$, the vector of coefficients of the OMLPSI (Cerón-Rojas et al., 2019). Thus, the OCMLPSI is more general and includes the multistage null phenotypic restricted index (Kempthorne and Nordskog, 1959; Xie and Xu, 1997) and the OMLPSI as particular cases (Cerón-Rojas and Crossa, 2018, Chapter 9).

The DCMLPSI Vector of Coefficients at Stage i

Let $I_{Di-1} = \mathbf{b}'_{i-1}\mathbf{x}_{i-1}$ and $I_{Di} = \mathbf{b}'_i\mathbf{x}_i$ be the DCMLPSIs at Stages $i-1$ and i, respectively. We shall obtain the DCMLPSI vector of coefficients at Stage i with the additional restriction that the covariance between the DCMLPSI values until Stage $i-1$ with the $I_{Di} = \mathbf{b}_i'\mathbf{x}_i$ values be null. Let $\mathbf{J}'_{i-1} = [I_{D1}\ I_{D2}\ \dots\ I_{D(i-1)}]$ be a vector of DCMLPSIs values until Stage $i-1$ such that the covariance between I_{Di} and $\mathbf{J}_{i-1}$ will be null. Xu and Muir (1992) and Xie and Xu (1997) showed that the covariance between I_{Di} and $\mathbf{J}_{i-1}$ is null when $\mathbf{B}'_{D(i-1)}\mathbf{Q}_{(i-1)i}\mathbf{b}_i = 0$, where

$$\mathbf{B}'_{D(i-1)} = \begin{bmatrix} \mathbf{b}'_1 & 0 & \ldots & 0 \\ \mathbf{b}'_1 & \mathbf{b}'_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}'_1 & \mathbf{b}'_2 & \ldots & \mathbf{b}'_{i-1} \end{bmatrix},$$

$\mathbf{b}'_{i-1} = [b_{(i-1)1}\ b_{(i-1)2}\ \ldots\ b_{(i-1)n_{(i-1)}}]$ is the DCMLPSI vector of coefficients at Stage $i-1$, $\mathbf{Q}_{(i-1)i}$ was defined in Eq. [1b] and $\mathbf{b}_i$ is the DCMLPSI vector of coefficients at Stage i. Thus, to obtain $\mathbf{b}_i$, we need to minimize the mean squared difference between $H = \mathbf{w}'\mathbf{g}$ and $I_{Di} = \mathbf{b}'_i\mathbf{x}_i$ {i.e., $E[(H - I_{Di})^2]$}, under the joint restrictions $\mathbf{M}'_i\mathbf{b}_i = 0$ and $\text{Cov}(I_{Di}, \mathbf{J}_{i-1}) = \mathbf{B}'_{D(i-1)}\mathbf{Q}_{(i-1)i}\mathbf{b}_i = 0$.

Let $\mathbf{S}_{i(i-1)} = \mathbf{Q}_{i(i-1)}\mathbf{B}_{D(i-1)}$ and $\mathbf{S}'_{i(i-1)} = \mathbf{S}_{(i-1)i} = \mathbf{B}'_{D(i-1)}\mathbf{Q}_{(i-1)i}$ be the transpose of matrix $\mathbf{S}_{i(i-1)}$ and assume that matrices $\mathbf{Q}_{ii}$, $\mathbf{Q}_{i(i-1)}$, $\mathbf{U}$, and $\mathbf{A}_i$ and vectors $\mathbf{d}$ and $\mathbf{w}$ are known. To minimize $E[(H - I_{Di})^2]$ under the restrictions $\mathbf{M}'_i\mathbf{b}_i = 0$ and $\mathbf{S}_{(i-1)i}\mathbf{b}_i = \mathbf{B}'_{D(i-1)}\mathbf{Q}_{(i-1)i}\mathbf{b}_i = 0$, it is necessary to minimize the function

$$f_D(\mathbf{b}_i, \mathbf{u}, \mathbf{v}) = \mathbf{b}'_i\mathbf{Q}_{ii}\mathbf{b}_i - 2\mathbf{w}\mathbf{A}'_i\mathbf{b}_i + 2\mathbf{u}'\mathbf{M}'_i\mathbf{b}_i + 2\mathbf{v}'\mathbf{S}_{(i-1)i}\mathbf{b}_i \quad [6]$$

with respect to the vector of coefficients $\mathbf{b}_i$ and the vector of Lagrange multipliers $\mathbf{u}' = [u_1\ u_2\ \ldots\ u_{r-1}]$ and $\mathbf{v}' = [v_1\ v_2\ \ldots\ v_{i-1}]$. The only difference between Eq. [4] and Eq. [6] is the term $2\mathbf{v}'\mathbf{S}_{(i-1)i\mathbf{b}i}$. The DCMLPSI vector of coefficients at Stage i is

$$\mathbf{b}_i = \mathbf{K}_{Di}\boldsymbol{\delta}_i \quad [7]$$

where $\boldsymbol{\delta}_i = \mathbf{Q}_{ii}^{-1}\mathbf{A}_i\mathbf{w}$, $\mathbf{Q}_{ii}^{-1}$, $\mathbf{A}_i$, and $\mathbf{w}$ were defined in Eq. [5]. In addition, $\mathbf{K}_{Di} = [\mathbf{I}_i - \mathbf{F}_{Di}]$, $\mathbf{F}_{Di} = \mathbf{Q}_{ii}^{-1}\mathbf{V}_i(\mathbf{V}'_i\mathbf{Q}_{11}^{-1}\mathbf{V}_i)^{-1}$, $\mathbf{V}_i = [\mathbf{M}_i\ \mathbf{S}_{i(i-1)}]$, $\mathbf{V}'_i = [\mathbf{M}'_i\ \mathbf{S}_{(i-1)i}]$, and $\mathbf{I}_i$ is an identity matrix of the same size as $\mathbf{Q}_{ii}$. Thus, the only difference between Eq. [5] and [7] is matrix $\mathbf{S}_{i(i-1)}$. When matrix $\mathbf{S}_{i(i-1)}$ is null, Eq. [5] and [7] are the same, as we would expect. If $\mathbf{D} = \mathbf{U}$, Eq. [6] and [7] impose null restrictions; in such cases, we shall have a multistage index similar to the Kempthorne and Nordskog (1959) index.

According to Eq. [5] and [7] results, matrices $\mathbf{K}_{Oi}$ and $\mathbf{K}_{Di}$ transform the OMLPSI vector of coefficients ($\boldsymbol{\delta}_i = \mathbf{Q}_{ii}^{-1}\mathbf{A}_i\mathbf{w}$) into the OCMLPSI and the DCMLPSI vectors of coefficients, respectively.

Maximized Selection Response at Stage i

It can be shown (Cerón-Rojas and Crossa, 2018, Chapter 9) that when we use Eq. [5] and [7] in Eq. [2], we obtain

$$R_{Oi} = k_{Oi}\sqrt{\boldsymbol{\beta}'_i\mathbf{Q}_{ii}\boldsymbol{\beta}_i} \quad [8a]$$

and

$$R_{Di} = k_{Di}\sqrt{\mathbf{b}'_i\mathbf{Q}_{ii}\mathbf{b}_i} \quad [8b]$$

which are the maximized OCMLPSI and DCMLPSI selection responses at Stage i, respectively. Although in Eq. [2] the selection response can take any value, in Eq. [8a] and [8b], R_{Oi} and R_{Di} give the maximum value of Eq. [2] for the OCMLPSI and DCMLPSI, respectively. In addition, in practice k_{Oi} and k_{Di} are obtained with a different method; therefore, their values are generally different (i.e., $k_{Oi} \neq k_{Di}$). In this work, we obtained the

k_{Oi} value in a two-stage breeding scheme according to Eq. [A6] (Appendix B), whereas we obtained the k_{Di} values according to the Xu and Muir (1992) method. We described a method to estimated Eq. [8a] and [8b] in Appendix B (Eq. [A1] to [A4b]).

The maximized total OCMLPSI and DCMLPSI selection responses for N stages can be written as $R_{t_O} = R_{O1} + R_{O2} + \ldots + R_{ON}$ and $R_{t_D} = R_{D1} + R_{D2} + \ldots + R_{DN}$, respectively.

Maximized Expected Genetic Gain per Trait at Stage i

Using Eq. [5] and [7] in Eq. [3], the OCMLPSI and DMLPSI expected genetic gains per trait at Stage i can be written as

$$\mathbf{E}_{Oi} = k_{Oi}\frac{\mathbf{A}'_i\boldsymbol{\beta}_i}{\sqrt{\boldsymbol{\beta}'_i\mathbf{Q}_{ii}\boldsymbol{\beta}_i}} \quad [9a]$$

and

$$\mathbf{E}_{Di} = k_{Di}\frac{\mathbf{A}'_i\mathbf{b}_i}{\sqrt{\mathbf{b}'_i\mathbf{Q}_{ii}\mathbf{b}_i}} \quad [9b]$$

respectively. We defined all the parameters of Eq. [9a] and [9b] earlier. The maximized total OCMLPSI and DMLPSI expected genetic gains per trait for N stages can be written as $\mathbf{E}_{t_O} = \mathbf{E}_{O1} + \mathbf{E}_{O2} + \ldots + \mathbf{E}_{ON}$ and $\mathbf{E}_{t_D} = \mathbf{E}_{D1} + \mathbf{E}_{D2} + \ldots + \mathbf{E}_{DN}$, respectively. We described a method to estimated Eq. [9a] and [9b] in Appendix B (Eq. [A1] to [A3], and Eq. [A5a] and [A5b]).

Efficiency when Predicting the Net Genetic Merit

According to Lande and Thompson (1990) and Moreau et al. (1998), the efficiency of the indices when predicting the net genetic merit, in percentage terms, is

$$\varnothing = 100(T - 1) \quad [10]$$

where $T = R_{Oi}/R_{Di}$, R_{Oi} denotes the OCMLPSI selection response and R_{Di} the DCMLPSI selection response. Therefore, when $\varnothing$ is null, the efficiency of both indices is the same; when $\varnothing > 0$, the efficiency of the OCMLPSI is higher than that of the DCMLPSI, and when $\varnothing < 0$, DCMLPSI efficiency is higher than OCMLPSI efficiency for predicting the net genetic merit.

An additional criterion for comparing the indices' efficiency is that the total selection response $R_t = R_1 + R_2$ of each index should be lower than or equal to the single-stage constrained index selection response ($R = k\sigma_1$), i.e., $R_t \leq R$ (see Cerón-Rojas et al., 2019, for details).

Adjusting the OCMLPSI Covariance Matrices at Stage 2

At Stage 2, the phenotypic covariance matrix is

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_2 \end{bmatrix} = \mathbf{Q}_{22}$$

whereas the genotypic covariance matrix is

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix} = \mathbf{A}_2 \quad (\text{Eq. [1a]})$$

These matrices are affected by prior selection on $I_1 = \boldsymbol{\beta}'\mathbf{x}_1$. It is thus necessary to adjust them to take into consideration the

$I_1 = \boldsymbol{\beta}'_1\mathbf{x}_1$ effects on them. According to Cochran (1951) and Cunningham (1975), both matrices can be adjusted as follows:

$$\mathbf{P}^* = \mathbf{P} - a\frac{\begin{bmatrix}\mathbf{P}_1 \\ \mathbf{P}_{21}\end{bmatrix}\boldsymbol{\beta}_1\boldsymbol{\beta}'_1\begin{bmatrix}\mathbf{P}_1 & \mathbf{P}_{21}\end{bmatrix}}{\boldsymbol{\beta}'_1\mathbf{P}_1\boldsymbol{\beta}_1} \qquad [11a]$$

and

$$\mathbf{G}^* = \mathbf{G} - a\frac{\mathbf{G}'_1\boldsymbol{\beta}_1\boldsymbol{\beta}'_1\mathbf{G}_1}{\boldsymbol{\beta}'_1\mathbf{P}_1\boldsymbol{\beta}_1} \qquad [11b]$$

where $\mathbf{P}^*$ and $\mathbf{G}^*$ are the adjusted matrix, $a = k_{O1}(k_{O1} - u)$, k_{O1} is the selection intensity at Stage 1, u is the truncation point when $I_1 = \boldsymbol{\beta}'_1\mathbf{x}_1$ is applied, $\mathbf{P}_1 = \mathrm{Var}(\mathbf{x}_1)$ and $\mathbf{G}_1 = \mathrm{Cov}(\mathbf{x}_1, \mathbf{g})$. Thus, the maximized OCMLPSI selection response (Eq. [8a]) and expected genetic gains (Eq. [9a]) at Stage 2 can be written as $R_{O2} = k_{O2}\sqrt{\boldsymbol{\beta}'_2\mathbf{P}^*\boldsymbol{\beta}_2}$ and $\mathbf{E}_{O2} = k_{O2}(\mathbf{G}^*\boldsymbol{\beta}_2/\sqrt{\boldsymbol{\beta}'_2\mathbf{P}^*\boldsymbol{\beta}_2})$, respectively.

Test of the OCMLPSI (DCMLPSI) Normality Assumption

Several authors (Shapiro and Wilk 1965; Mardia 1980; Mohd-Razali and Bee-Wah 2011; Rani Das and Rahmatullah Imon, 2016) have given details of how to perform a normality test procedure on a dataset and many statistical packages provide graphs and normality tests.

We corroborated the OCMLPSI (DCMLPSI) normality assumption at Stage 2 with a simulated dataset using a graphical method (histograms) and analytical test procedures (the Shapiro–Wilk and Kolmogorov–Smirnov normality test). The corroboration procedure was as follows. In a two-stage context, let $p = q_1q_2$ be the fixed total proportion retained, where q_1 and q_2 denote the proportion selected at Stage 1 and 2, respectively, and let n be the size of the simulated dataset at Stage 1; then, nq_1 will be the size of the selected individuals at Stage 1. We used the information of nq_1 individuals at Stage 2 to construct graphs and statistical tests to corroborate the OCMLPSI (DCMLPSI) normality assumption.

Materials

Real Dataset

The number of genotypes in this real data set was 3330 and the vector of economic weights ($\mathbf{w}$) was $\mathbf{w}' = [19.54 \quad -3.56 \quad 17.01 \quad -2.51]$. This dataset comes from a commercial egg poultry line (Akbar et al., 1984) and we used it to illustrate the indices' theoretical results obtained in this work. The estimated phenotypic ($\hat{\mathbf{P}}$) and genotypic ($\hat{\mathbf{C}}$) covariance matrices among the rate of lay (RL, number of eggs), age at sexual maturity (SM, d), egg weight (EW, kg), and body weight (BW, kg) were

$$\hat{\mathbf{P}} = \begin{bmatrix} 240.57 & -95.62 & 2.06 & 54.40 \\ -95.62 & 167.20 & 4.58 & 15.36 \\ 2.06 & 4.58 & 22.80 & 37.20 \\ 54.40 & 15.36 & 37.20 & 516.11 \end{bmatrix}$$

and

$$\hat{\mathbf{C}} = \begin{bmatrix} 29.86 & -17.90 & -4.13 & -1.75 \\ -17.90 & 18.56 & 1.49 & -4.88 \\ -4.13 & 1.49 & 9.24 & 16.66 \\ -1.75 & -4.88 & 16.66 & 179.73 \end{bmatrix}$$

The total proportions (p) of retained values for this dataset were $p = 0.05, 0.10, 0.20,$ and 0.30 for both indices.

For illustration purposes only, at Stage 1, we selected RL, SM, and EW, where RL and SM were constrained by the vector of predetermined restrictions $\mathbf{d}' = [3 \quad -1]$ and matrices

$$\mathbf{U}' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

and $\mathbf{D}' = [-1 \quad -3]$ for both stages, whereas at Stage 2, we selected trait BW only. The vectors of records at Stages 1 and 2 were $\mathbf{x}'_1 = [\mathrm{RL} \quad \mathrm{SM} \quad \mathrm{EW}]$ and $\mathbf{y}' = [\mathbf{x}'_1 \quad \mathbf{x}'_2]$, respectively, where $\mathbf{x}_2 = \mathrm{BW}$. At Stage 1, the estimated phenotypic ($\hat{\mathbf{Q}}_{11}$) and genotypic ($\hat{\mathbf{A}}_1$) covariance matrices were

$$\hat{\mathbf{Q}}_{11} = \begin{bmatrix} 240.57 & -95.62 & 2.06 \\ -95.62 & 167.20 & 4.58 \\ 2.06 & 4.58 & 22.80 \end{bmatrix}$$

and

$$\hat{\mathbf{A}}_1 = \begin{bmatrix} 29.86 & -17.90 & -4.13 & -1.75 \\ -17.90 & 18.56 & 1.49 & -4.88 \\ -4.13 & 1.49 & 9.24 & 16.66 \end{bmatrix}$$

respectively, whereas the estimated covariance matrix of the traits at Stage 1 with traits at Stage 2 ($\hat{\mathbf{Q}}_{12}$) was

$$\hat{\mathbf{Q}}_{12} = \begin{bmatrix} 240.57 & -95.62 & 2.06 & 54.40 \\ -95.62 & 167.20 & 4.58 & 15.36 \\ 2.06 & 4.58 & 22.88 & 37.20 \end{bmatrix}$$

At Stage 2, $\hat{\mathbf{Q}}_{22} = \hat{\mathbf{P}}$ and $\hat{\mathbf{A}}_2 = \hat{\mathbf{C}}$.

Simulated Datasets

These datasets are available in the *Application of a Genomics Selection Index to Real and Simulated Data* repository, at http://hdl.handle.net/11529/10199. They were simulated by Ceron-Rojas et al. (2015) with QU-GENE software (Podlich and Cooper, 1998) using 2500 molecular markers and 315 quantitative trait loci (QTLs) for eight phenotypic selection cycles (C0–C7), each with four traits (T_1, T_2, T_3, and T_4), 500 genotypes and four replicates for each genotype. The authors distributed the markers uniformly across 10 chromosomes and the QTLs randomly across the 10 chromosomes to simulate maize (*Zea mays* L.) populations. A different number of QTLs affected each of the four traits: 300, 100, 60, and 40, respectively. The common QTLs affecting the traits generated genotypic correlations of -0.5, 0.4, 0.3, -0.3, -0.2, and 0.1 between T_1 and T_2,

T_1, and T_3, T_1 and T_4, T_2, and T_3, T_2 and T_4, T_3 and T_4, respectively. The economic weights for T_1, T_2, T_3, and T_4 were 1, −1, 1, and 1, respectively.

We used four phenotypic selection cycles (C1–C4) with p = 0.01, 0.10, and 0.30 in each cycle. At Stage 1 we selected T_1, T_2, and T_3, where T_1 and T_2 were constrained with vector $\mathbf{d}' = $ [5 −2] and matrices

$$\mathbf{U}' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

and $\mathbf{D}' = $ [−2 −5] for both stages. At Stage 2 we selected only trait T_4; thus, the vector of observations at Stage 1 was $\mathbf{x}'_1 = $ [T_1 T_2 T_3] and at Stage 2, $\mathbf{y}' = $ [$\mathbf{x}'_1$ $\mathbf{x}'_2$], where $\mathbf{x}_2 = T_4$.

RESULTS
Real Data
Truncation Points, Proportion Retained, and Selection Intensities for Two Stages

Figure 1 shows the relationship among the truncation points (u_1 and u_2), the total proportion retained ($p = q_1 q_2$) and the heights of the ordinate of the normal curve: $z(u_1) = e^{-0.5 u_1^2} / \sqrt{2\pi}$ and $z(u_2) = e^{-0.5 u_2^2} / \sqrt{2\pi}$. We found the OCMLPSI selection intensity for Stages 1 [$k_1 = z(u_1)/q_1$] and 2 [$k_2 = z(u_2)/q_2$] according to Eq. [A6] (Appendix B) as follows. For a fixed value of $p = q_1 q_2$ (e.g., $p = $ 0.05), we used an iterative process with an R code. By successively changing the possible values of q_1 ($q_2 = p/q_1$), u_1, and u_2, we found the maximum value of the estimated total OCMLPSI (DCMLPSI) selection response, $\hat{R}_t = \hat{R}_1 + \hat{R}_2$ (Fig. 2). For example, for the real dataset and $p = $ 0.05, the estimated total OCMLPSI selection response was $\hat{R}_t = \hat{R}_1 + \hat{R}_2 = 25.592 k_1 + 26.580 k_2 = 69.75$, where $\hat{R}_1 = k_1 \hat{\sigma}_{I_1} = 33.265$ and $\hat{R}_2 = k_2 \hat{\sigma}_{I_2} = 36.481$ were the estimated selection responses at each stage (Table 1), whereas $\hat{\sigma}_{I_1} = 25.592$ and $\hat{\sigma}_{I_2} = 26.579$ were the estimated standard deviations of the variance of $\hat{I}_1$ and $\hat{I}_2$ for Stages 1 and 2, respectively. Thus, for this dataset, the values of the truncation points ($u_1 = 0.710$ and $u_2 = 0.81$), proportions retained ($q_1 = 0.24$ and $q_2 = 0.21$) and selection intensity ($k_1 = 1.30$ and $k_2 = 1.37$), at both stages, were those associated with the maximum estimated total OCMLPSI selection response $\hat{R}_t = 69.75$ value. Table 1, presents additional truncation points, proportions retained, selection intensities for $p = q_1 q_2 = $ 0.10, 0.20, and 0.30, associated to the OCMLPSI and DCMLPSI selection responses.

Estimated OCMLPSI Selection Response for Two Stages

In the one-stage case, the selection intensity for $p = $ 0.05 was $k = 2.063$, and the SCLPSI selection response was $\hat{R}$ = 71.66 (see Cerón-Rojas and Crossa, 2018, Chapter 3, for details). According to the results detailed in the paragraph
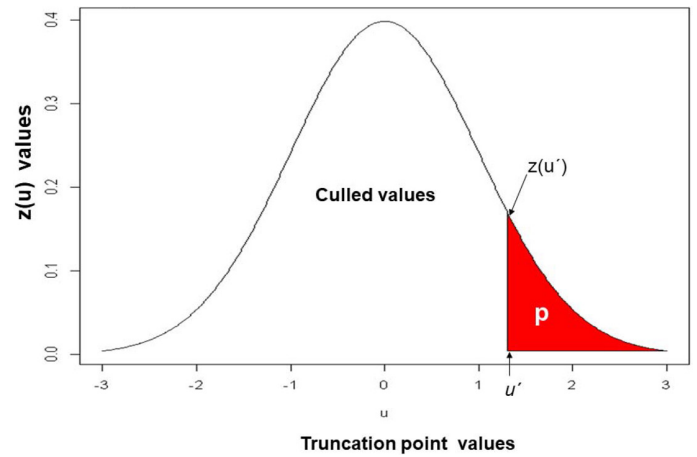

Fig. 1. Theoretical relationship between the truncation points (u), the proportion retained (p), and the density values [$z(u)$] of the truncation points (after Cerón-Rojas et al., 2019).

above and to Young (1964) and Saxton (1983), the maximum estimated total OCMLPSI selection response ($\hat{R}_t = 69.75$) value should be lower than or equal to the estimated SCLPSI response ($\hat{R} = 71.66$). For this dataset, $\hat{R}_t = 69.75$ explained 97.33% of the $\hat{R} = 71.66$ value. That is, $\hat{R}_t = 69.75$ and $\hat{R} = 71.66$ were very similar.

In Table 1, we present additional maximum estimated OCMLPSI selection response values for $p = q_1 q_2 = $ 0.10, 0.20, and 0.30. For each of the latter three p values, the maximum estimated total OCMLPSI selection responses explained 97.44, 97.68, and 97.96%, respectively, of the SCLPSI selection response values. Thus, for this real dataset, the estimated total OCMLPSI selection response and the estimated SCLPSI selection response were very similar, as we would expect.

Estimated DCMLPSI Selection Response for Two Stages

For both stages, the estimated DCMLPSI vectors of coefficients (Eq. [7]) were $\hat{\mathbf{b}}'_1 = $ [1.809 1.132 1.065] and $\hat{\mathbf{b}}'_2 = $ [0.349 −0.262 1.743 −1.076]. Because at Stage 1 the restriction matrix $\mathbf{S}_{(i-1)i}$ was null [$\mathbf{S}_{(i-1)i} = 0$], the estimated DCMLPSI vector of coefficients was the same as the OCMLPSI vector of coefficients (i.e., $\hat{\mathbf{b}}_1 = \hat{\boldsymbol{\beta}}'_1$; Eq. [5] and [7]); however, at Stage 2, $\mathbf{S}_{(i-1)i} \neq 0$ and $\hat{\mathbf{b}}_2 \neq \hat{\boldsymbol{\beta}}'_2$.

For $p = $ 0.05, the selection intensities obtained with the Xu and Muir (1992, Eq. [19]) method were $k_1 = 1.42$ and $k_2 = 1.25$ at Stages 1 and 2, respectively, from where the estimated maximized selection responses for both stages were $\hat{R}_1 = 39.29$ and $\hat{R}_2 = 29.40$ (Appendix A, Eq. [A4b]), whereas $\hat{R}_t = \hat{R}_1 + \hat{R}_2 = 65.69$ was the total estimated selection response. This means that $\hat{R}_t = 65.69$ explained 91.67% of the estimated SCLPSI selection response ($\hat{R} = 71.66$) value.

Table 1 presents additional maximum estimated DCMLPSI selection response values when $p = q_1 q_2 = $ 0.10, 0.20, and 0.30. For 0.10, 0.20 and 0.30, the estimated total
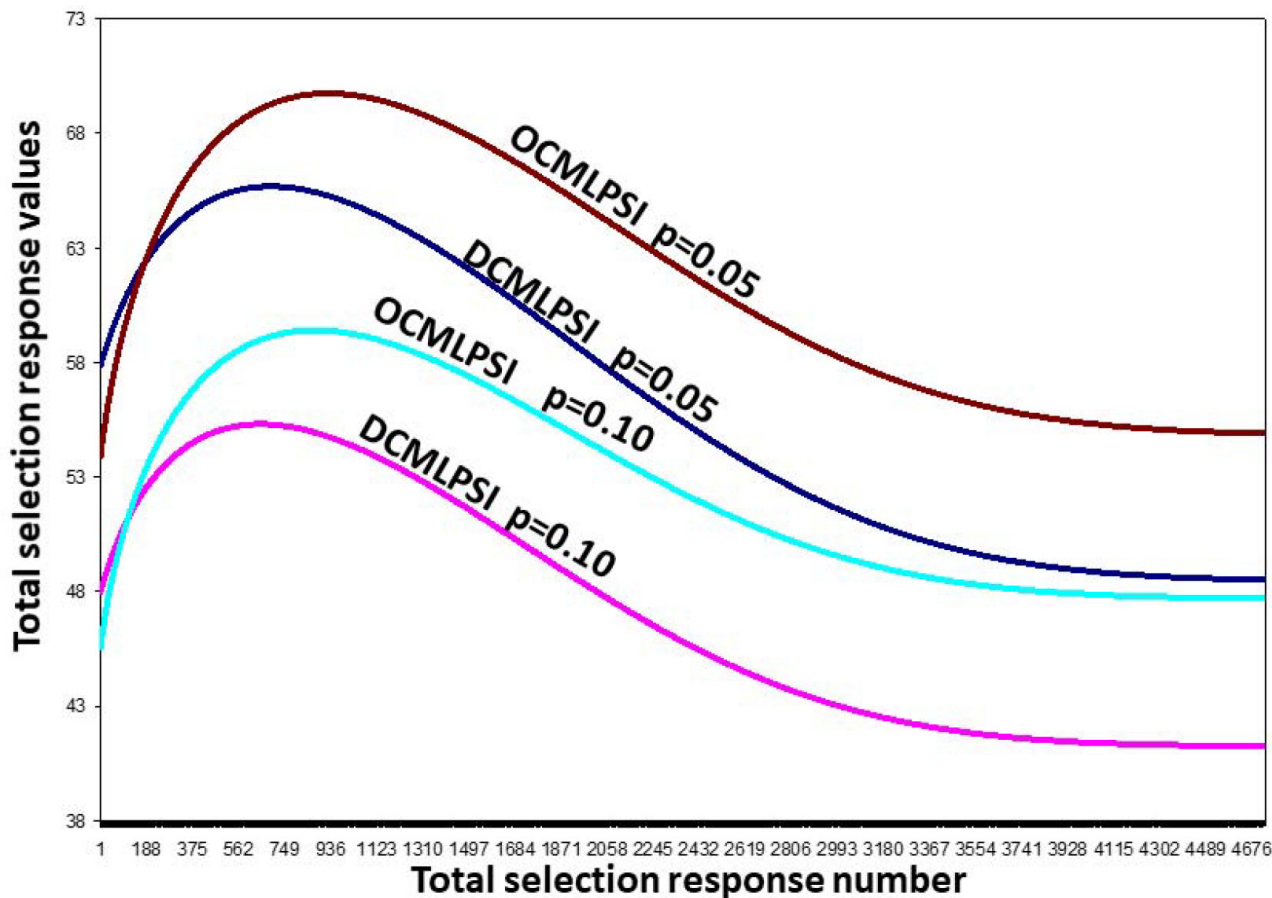
Fig. 2. Distribution of the total estimated optimum and decorrelated constrained multistage linear phenotypic selection index (OCMLPSI and DCMLPSI, respectively) selection response values for a real dataset, and the fixed total proportion retained (p) = 0.05 and 0.10.

selection response explained 90.72, 89.40, and 88.30%, respectively, of the estimated SCLPSI selection response.

The results of the last two subsections indicate that the average of the estimated total DCMLPSI and OCMLPSI selection responses explained 90 and 97.60%, respectively, of the average estimated SCLPSI selection response for all p values. This means that the average of the estimated total OCMLPSI selection response was 7.60% closer to the estimated SCLPSI selection response than the average of the estimated total DCMLPSI selection response. We explain the loss of DCMLPSI efficiency by noting that when DCMLPSI obtained its vector of coefficients, it incorporated

an additional restriction, which made the DCMLPSI values independent at different stages. Xu and Muir (1992) and Xie and Xu (1997) indicated that the loss of efficiency is justified because their method for obtaining the selection intensities and total responses gives the breeder the opportunity to implement an unlimited number of selection stages, which otherwise would be very difficult or impossible to do.

Estimated OCMLPSI Expected Genetic Gain per Trait for Two Stages

Let p = 0.05 (k_1 = 1.30 and k_2 = 1.37); then, the estimated OCMLPSI expected genetic gains per trait (Appendix A,

Table 1. Real data for total proportion (p) retained, estimated optimum and decorrelated constrained multistage linear phenotypic selection index (OCMLPSI and DCMLPSI, respectively) truncation points (u_1 and u_2), proportions retained (q_1 and q_2), selection intensities (k_1 and k_2), and selection response ($\hat{R}_1$, $\hat{R}_2$, and $\hat{R}_t = \hat{R}_1 + \hat{R}_2$) for Stages 1 and 2. Values of $\hat{R}$ correspond to the one-stage estimated constrained linear phenotypic selection index selection response.

Index	p	u_1	u_2	q_1	q_2	k_1	k_2	$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_t$	$\hat{R}$
OCMLPSI	0.05	0.71	0.81	0.24	0.21	1.30	1.37	33.26	36.48	69.75	71.66
	0.10	0.41	0.55	0.34	0.29	1.07	1.18	27.45	31.95	59.40	60.96
	0.20	0.03	0.23	0.49	0.41	0.81	0.95	20.85	26.65	47.50	48.63
	0.30	−0.25	0.00	0.60	0.50	0.64	0.80	16.48	22.96	39.44	40.26
DCMLPSI	0.05	0.86	0.65	0.19	0.26	1.42	1.25	36.29	29.40	65.69	71.66
	0.10	0.58	0.37	0.28	0.35	1.20	1.05	30.65	24.66	55.30	60.96
	0.20	0.23	0.03	0.41	0.49	0.95	0.82	24.23	19.24	43.47	48.63
	0.30	−0.03	−0.22	0.51	0.59	0.78	0.67	19.91	15.64	35.55	40.26

Eq. [A5a]) for both stages were $\hat{\mathbf{E}}_1' = [1.49 \quad -0.50$ $0.21 \quad 0.46]$ and $\hat{\mathbf{E}}_2' = [1.92 \; -0.64 \; 0.01 \; -8.02]$, while $\hat{\mathbf{E}}_t' = \hat{\mathbf{E}}_1' + \hat{\mathbf{E}}_2' = [3.41 \quad -1.14 \quad 0.21 \quad -7.56]$ was the total estimated expected genetic gain per trait. Each $\hat{\mathbf{E}}_t'$ value was associated with the mean of traits rate of lay (RL, number of eggs), age at sexual maturity (SM, days), egg weight (EW, kg) and body weight (BW, kg). We constrained traits RL and SM by vector $\mathbf{d}' = [3 \quad -1]$ values. This means that the $\hat{\mathbf{E}}_t'$ values associated with traits RL and SM (3.41 and −1.14, respectively) overestimated the $\mathbf{d}' = [3 \quad -1]$ values by 10.33 and 14%, respectively.

Table 2 presents additional estimated expected genetic gains per traits RL, SM, EW, and BW for both stages and $p = 0.10$, 0.20, and 0.30. For these last three p values, the estimated total OCMLPSI expected genetic gains per traits RL and SM explained 95, 74, and 59% of each $\mathbf{d}' = [3 \quad -1]$ value, respectively. That is, the accuracy of the estimated OCMLPSI expected genetic gains per trait decreased when the p values increased from 0.10 to 0.30. For this real dataset, the optimum expected genetic gain per trait efficiency occurred when $p = 0.10$.

Estimated DCMLPSI Expected Genetic Gain per Trait for Two Stages

For $p = 0.05$ ($k_1 = 1.42$ and $k_2 = 1.25$), the estimated DCMLPSI expected genetic gains per trait for both stages were $\hat{\mathbf{E}}_1' = [1.63 \quad -0.54 \quad 0.22 \quad 0.50]$ and $\hat{\mathbf{E}}_2' = [0.52$ $-0.17 \quad -0.19 \quad -8.72]$, whereas $\hat{\mathbf{E}}_t' = \hat{\mathbf{E}}_1' + \hat{\mathbf{E}}_2' = [2.15$ $-0.72 \quad 0.03 \quad -8.22]$ was the total estimated expected genetic gain per trait. Each $\hat{\mathbf{E}}_t'$ value is associated with traits RL (rate of lay, number of eggs), SM (age at sexual maturity, days), EW (egg weight, kg), and BW (body weight, kg), and traits RL and SM were constrained by vector $\mathbf{d}' = [3 \quad -1]$ values. The $\hat{\mathbf{E}}_t'$ values associated with RL and SM (2.15 and −0.72, respectively) explained only 71.67 and 72% of each $\mathbf{d}' = [3 \quad -1]$ value, respectively.

Table 2 presents additional estimated expected genetic gains per traits RL, SM, EW, and BW for both stages and $p = 0.10$, 0.20, and 0.30. For 0.10, 0.20, and 0.30, the estimated total expected genetic gains per traits RL

and SM explained 60.33, 47.67, and 39% of each $\mathbf{d}' = [3 \quad -1]$ value, respectively. Thus, for this dataset, the estimated expected genetic gains per trait underestimated the $\mathbf{d}' = [3 \quad -1]$ values. We explained the loss of DCMLPSI accuracy, noting that when the DCMLPSI obtained its vector of coefficients, it incorporated an additional restriction to make the DCMLPSI values independent among stages. The average estimated DCMLPSI expected genetic gain per trait efficiency was 54.67%, whereas the average estimated OCMLPSI expected genetic gain per trait efficiency was 85% for all p values. Thus, the average of the estimated OCMLPSI accuracy associated with $\mathbf{d}' = [3 \quad -1]$ values was 35% higher than the average of the estimated DCMLPSI accuracy associated with $\mathbf{d}' = [3 \quad -1]$ for the real dataset.

The results of the above four subsections indicate that the accuracy of both indices was higher when they predicted the selection response than when they estimated the expected genetic gain per trait. However, for the real data, the efficiency of the OCMLPSI when predicting the selection response and estimating the expected genetic gain per trait was higher than the DMLPSI efficiency when predicting the selection response and estimating the expected genetic gain per trait.

OCMLPSI Efficiency vs. DCMLPSI Efficiency to Predict the Net Genetic Merit

Equation [10] is a tool for determining OCMLPSI efficiency vs. DCMLPSI efficiency when predicting the net genetic merit in percentage terms. The estimated average OCMLPSI efficiency to predict the net genetic merit in percentage terms is $100(97.604/90.019 - 1) = 8.426\%$, where 97.604 and 90.019 are the average of the estimated total OCMLPSI and DCMLPSI selection responses (Table 1) for all p values, respectively, and 8.426% is OCMLPSI efficiency with respect to DCMLPSI efficiency, in percentage terms, to predict the net genetic merit. Thus, for the Akbar et al. (1984) real dataset, the estimated average OCMLPSI efficiency was 8.426% higher than the estimated average DCMLPSI efficiency for predicting the net genetic merit.

Table 2. Total proportion (p) retained; estimated optimum and decorrelated constrained multistage linear phenotypic selection index (OCMLPSI and DCMLPSI, respectively) expected genetic gains per trait for four real traits: rate of lay (RL), age at sexual maturity (SM), egg weight (EW), and body weight (BW). Traits RL and SM were constrained with vector d′ (a vector of constraints or predetermined proportional gains per trait imposed by the breeder) = [3 −1] values.

		Stage 1				Stage 2				Total			
Index	p	RL	SM	EW	BW	RL	SM	EW	BW	RL	SM	EW	BW
		no. eggs	d	— kg —		no. eggs	d	— kg —		no. eggs	d	— kg —	
OCMLPSI	0.05	1.49	−0.50	0.21	0.46	1.92	−0.64	0.01	−8.02	3.41	−1.14	0.21	−7.56
	0.10	1.23	−0.41	0.17	0.38	1.62	−0.54	0.01	−6.72	2.85	−0.95	0.18	−6.34
	0.20	0.93	−0.31	0.13	0.29	1.27	−0.42	0.01	−5.26	2.21	−0.74	0.14	−4.97
	0.30	0.74	−0.25	0.10	0.23	1.04	−0.35	0.01	−4.30	1.78	−0.59	0.11	−4.07
DCMLPSI	0.05	1.63	−0.54	0.22	0.50	0.52	−0.17	−0.19	−8.72	2.15	−0.72	0.03	−8.22
	0.10	1.37	−0.46	0.19	0.42	0.44	−0.15	−0.16	−7.32	1.81	−0.60	0.03	−6.89
	0.20	1.09	−0.36	0.15	0.34	0.34	−0.11	−0.13	−5.71	1.43	−0.48	0.02	−5.37
	0.30	0.89	−0.30	0.12	0.28	0.28	−0.09	−0.10	−4.64	1.17	−0.39	0.02	−4.36

Simulated Data
Estimated OCMLPSI and DCMLPSI Selection Response

For $p = q_1q_2 = 0.01$, 0.10, and 0.30, Table 3 presents the estimated OCMLPSI and DCMLPSI responses ($\hat{R}_1$, $\hat{R}_2$, $\hat{R}_t = \hat{R}_1 + \hat{R}_2$) and SCLPSI responses ($\hat{R}_{0.01}$, $\hat{R}_{0.10}$, $\hat{R}_{0.30}$) for four simulated selection cycles in a two-stage breeding selection scheme. For $p = 0.01$, the average of the estimated total OCMLPSI selection responses (22.79) explained 99.80% of the average of the SCLPSI selection responses (22.84), whereas for $p = 0.10$ and 0.30, the average of the estimated total OCMLPSI selection responses (15.13 and 10.11, respectively) explained 100.60 and 101.81% of the average of the SCMLPSI selection response (15.04 and 9.93, respectively). Thus, for this dataset, the OCMLPSI and SCLPSI results were equivalent for all p values.

For $p = 0.01$, the average of the estimated total DCMLPSI selection responses (21.84) explained 95.62% of the average of the SCMLPSI selection responses (22.84), whereas for $p = 0.10$ and 0.30, the average of the estimated total DCMLPSI selection responses (14.43 and 9.49, respectively) explained 95.94 and 95.57% of the average of the SCMLPSI selection responses (15.04 and 9.93, respectively).

The results in this section indicate that although the average of the total OCMLPSI selection response, for all p values, overestimated the average of the SCLPSI by 0.73%, the average of the total DCMLPSI selection response, for all p values, underestimated the average of the SCLPSI by 4.30%. Thus, for this simulated dataset, the OCMLPSI was the best predictor of the net genetic merit, and its accuracy when predicting the selection response was higher than the DMLPSI accuracy for predicting the selection response.

Estimated OMLPSI and DMLPSI Expected Genetic Gains per Trait

Table 4 presents the estimated OCMLPSI and DCMLPSI expected genetic gains per trait ($\hat{\mathbf{E}}'_1$, $\hat{\mathbf{E}}'_2$, and $\hat{\mathbf{E}}'_t = \hat{\mathbf{E}}'_1 + \hat{\mathbf{E}}'_2$)

for four simulated selection cycles and $p = q_1q_2 = 0.30$ in a two-stage context. Each $\hat{\mathbf{E}}'_t$ value was associated with the mean values of traits T_1, T_2, T_3, and T_4. In addition, in both indices, traits T_1 and T_2 were constrained by vector $\mathbf{d}' = [5 \quad -2]$ values. The average of the estimated total OCMLPSI $\hat{\mathbf{E}}'_t$ values associated with traits T_1 and T_2 (5.76 and −2.30, respectively) overestimated the $\mathbf{d}' = [5 \quad -2]$ values by 15.20%. However, the average of the estimated total DCMLPSI $\hat{\mathbf{E}}'_t$ values associated with traits T_1 and T_2 (5.05 and −2.02, respectively) overestimated the $\mathbf{d}' = [5 \quad -2]$ values by only 1.0%. Nevertheless, note that at Stage 2, the averages of the estimated total DCMLPSI expected genetic gains per trait associated with traits T_1, T_2, and T_3 (0.02, −0.01, and 0.0, respectively) were practically null. This means that DCMLPSI efficiency occurred at Stage 1, when restriction matrix $\mathbf{S}_{(i-1)i}$ was null [$\mathbf{S}_{(i-1)i} = 0$] and $\hat{\mathbf{b}}_1 = \hat{\boldsymbol{\beta}}'_1$. Thus, for this dataset, the average of the estimated total DCMLPSI expected genetic gains per trait was more efficient for predicting the $\mathbf{d}' = [5 \quad -2]$ values than the OCMLPSI, but the highest DCMLPSI efficiency occurred at Stage 1, when $\hat{\mathbf{b}}_1 = \hat{\boldsymbol{\beta}}'_1$ and the estimated standard deviations of OCMLPSI and DCMLPSI values were the same.

We also estimated the total expected genetic gains per trait of both indices for $p = 0.01$, 0.10, and 0.20 (data not shown); however, in all cases, those values were higher than the $\mathbf{d}' = [5 \quad -2]$ values. For example, for $p = 0.10$, the averages of the estimated total OCMLPSI and DCMLPSI expected genetic gains per trait associated with $\mathbf{d}' = [5 \quad -2]$ were 8.78 and −3.51, and 7.58 and −3.03, respectively.

The difference between the OCMLPSI and DCMLPSI expected genetic gains per trait is due to the different number of genotypes used to estimate the parameters. That is, in the real dataset, the number of genotypes was 3330, but in the simulated data, the number of genotypes was only 500, which represents only 15% of the size of the genotypes used in the real dataset to estimate the parameters of the

Table 3. Simulated data for total proportion retained (p) = q_1q_2 = 0.01, 0.10, and 0.30, and estimated optimum and decorrelated constrained multistage linear phenotypic selection indices (OCMLPSI and DCMLPSI, respectively) responses ($\hat{R}_1$, $\hat{R}_2$, and $\hat{R}_t = \hat{R}_1 + \hat{R}_2$) and single-stage constrained linear phenotypic selection index (SCLPSI) responses ($\hat{R}_{0.01}$, $\hat{R}_{0.10}$, and $\hat{R}_{0.30}$) for four simulated selection cycles in a two-stage breeding scheme.

| | | OCMLPSI | | | | | | | | | SCLPSI | | |
| | | p = 0.01 | | | p = 0.10 | | | p = 0.30 | | | | | |
Index	Cycle	$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_t$	$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_t$	$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_t$	$\hat{R}_{0.01}$	$\hat{R}_{0.10}$	$\hat{R}_{0.30}$
OCMLPSI	1	21.59	3.16	24.75	13.87	2.57	16.44	8.75	2.24	10.99	24.77	16.31	10.77
	2	20.39	3.09	23.47	12.99	2.6	15.59	8.18	2.24	10.42	23.54	15.50	10.24
	3	17.75	3.48	21.23	11.28	2.81	14.09	7.09	2.32	9.41	21.39	14.08	9.30
	4	18.95	2.73	21.68	12.21	2.19	14.4	7.70	1.93	9.63	21.68	14.28	9.43
	Avg.	19.67	3.12	22.79	12.59	2.54	15.13	7.93	2.18	10.11	22.84	15.04	9.93
DCMLPSI	1	21.59	2.08	23.68	15.25	0.47	15.72	10.15	0.20	10.35	24.77	16.31	10.77
	2	20.39	2.12	22.51	14.40	0.48	14.88	9.58	0.20	9.78	23.54	15.50	10.24
	3	18.17	2.27	20.44	12.57	0.80	13.36	8.43	0.33	8.76	21.39	14.08	9.30
	4	18.95	1.77	20.72	13.38	0.40	13.78	8.91	0.17	9.07	21.68	14.28	9.43
	Avg.	19.78	2.06	21.84	13.90	0.54	14.43	9.27	0.22	9.49	22.84	15.04	9.93

indices. This means that the number of genotypes used to estimate the indices' parameters was an important factor for both indices in the real and simulated data.

The results of the real and simulated datasets indicated that the OCMLPSI is the most efficient index for predicting the net genetic merit, and its accuracy when predicting the selection response and estimating the expected genetic gain per trait was higher than DCMLPSI accuracy when predicting the selection response and estimating the expected genetic gain per trait.

Normality Test for the Estimated OCMLPSI and DCMLPSI Values at Stage 2

We used the simulated dataset in Cycle 2 to test the normality assumption of the estimated OCMLPSI and DCMLPSI values at Sage 2. In Cycle 1, the number of genotypes was 500. For $p = q_1 q_2 = 0.05$ and 0.30, the q_1 values for OCMLPSI were 0.22 and 0.55, whereas those values for DCMLPSI were 0.06 and 0.31, respectively. Then, at Stage 2, $(0.2)(500) = 110$ and $(0.55)(500) = 270$ were the number of genotypes for OCMLPSI, whereas for DCMLPSI, the number of genotypes were $(0.06)(500) = 30$ and $(0.31)(500) = 155$. We used these last numbers of genotypes to construct histograms (Fig. 3) of the estimated OCMLPSI and DCMLPSI values at Stage 2.

According to the histograms constructed for the estimated OCMLPSI values, when the number of genotypes changed from 110 (Fig. 3a) to 270 (Fig. 3b), the estimated OCMLPSI values were closer to the normal distribution. The same was true for the estimated DCMLPSI values (Fig. 3c and 3d).

We describe now the Shapiro–Wilk and Kolmogorov–Smirnov normality test results of the estimated OCMLPSI and DCMLPSI values at Stage 2 (Cycle 2) when the number of genotypes was 110 and 270 for OCMLPSI, and 30 and 155 for DCMLPSI. With the simulated dataset, we tested the null hypothesis that the estimated OCMLPSI and DCMLPSI values at Stage 2 have normal distribution.

The statistical value of the Shapiro–Wilk test should be close to 1.0 to accept the null hypothesis, whereas the statistic value of the Kolmogorov–Smirnov test should be close to 0.0 to accept the null hypothesis (Rani Das and Rahmatullah Imon, 2016). In the present case, for the values associated with OCMLPSI (110 and 270), the statistic values of the Shapiro–Wilk were 0.958 and 0.989, whereas the statistic values of the Kolmogorov–Smirnov were 0.080 and 0.044, respectively. Thus, we believe that for the estimated OCMLPSI values, the null hypothesis was true. In a similar manner, for the values associated with DCMLGSI (30 and 155), the statistic values of the Shapiro–Wilk were 0.967 and 0.991, whereas the statistic values of the Kolmogorov–Smirnov were 0.094 and 0.029, respectively. We again accept that the estimated DCMLPSI values approach the normal distribution.

DISCUSSION
Criteria Used to Evaluate the Relative Efficiency of the Indices

A criterion used to evaluate OCMLPSI efficiency vs. DCMLPSI efficiency when predicting the net genetic merit was that the estimated total OCMLPSI and DCMLPSI selection response must be lower than or equal to the single-stage estimated OCLPSI selection response. Additional criteria were the ratio of the OCMLPSI selection response over the DCMLPSI selection response and the estimated expected genetic gain per trait or multitrait selection response. The estimated total selection response of both indices predicted the mean value of the net genetic merit in the progeny population, whereas the estimated expected genetic gain values indicated how close the estimated mean values of the traits are to the predetermined proportional gains (or constraints) imposed by the breeder in each selection cycle. Both parameters are good criteria for comparing the efficiency of the indices, depending on the method used to estimate the vector of coefficients of each index.

Table 4. Estimated optimum and decorrelated constrained multistage linear phenotypic selection indices (OCMLPSI and DCMLPSI, respectively) expected genetic gains per trait ($\hat{E}'_1$, $\hat{E}'_2$, and $\hat{E}'_t = \hat{E}'_1 + \hat{E}'_2$) for Stages 1 and 2 in four simulated selection cycles with total proportion retained (p) = $q_1 q_2$ = 0.30. Traits T_1 and T_2 were constrained with vector d' (a vector of constraints or predetermined proportional gains per trait imposed by the breeder) = [5 −2] values on the two indices.

| Index	Cycle	Stage 1 ($\hat{E}'_1$)				Stage 2 ($\hat{E}'_2$)				$\hat{E}'_t = \hat{E}'_1 + \hat{E}'_2$			
		T_1	T_2	T_3	T_4	T_1	T_2	T_3	T_4	T_1	T_2	T_3	T_4
OCMLPSI	1	4.55	−1.82	2.08	0.31	1.49	−0.59	0.66	0.54	6.03	−2.41	2.73	0.85
	2	4.30	−1.72	1.87	0.29	1.39	−0.56	0.60	0.59	5.69	−2.28	2.47	0.88
	3	3.87	−1.55	1.58	0.09	1.48	−0.59	0.57	0.64	5.36	−2.14	2.15	0.73
	4	4.50	−1.80	1.26	0.14	1.44	−0.58	0.40	0.43	5.94	−2.38	1.67	0.56
	Avg.	4.31	−1.72	1.70	0.21	1.45	−0.58	0.56	0.55	5.76	−2.30	2.25	0.76
DCMLPSI	1	5.28	−2.11	2.41	0.35	0.01	−0.01	0.00	0.18	5.29	−2.12	2.41	0.53
	2	5.04	−2.02	2.19	0.34	0.01	0.00	0.00	0.18	5.05	−2.02	2.20	0.52
	3	4.60	−1.84	1.87	0.11	0.04	−0.01	0.00	0.28	4.64	−1.86	1.87	0.39
	4	5.21	−2.08	1.46	0.16	0.00	0.00	0.00	0.16	5.21	−2.08	1.46	0.32
	Avg.	5.03	−2.01	1.98	0.24	0.02	−0.01	0.00	0.20	5.05	−2.02	1.98	0.44
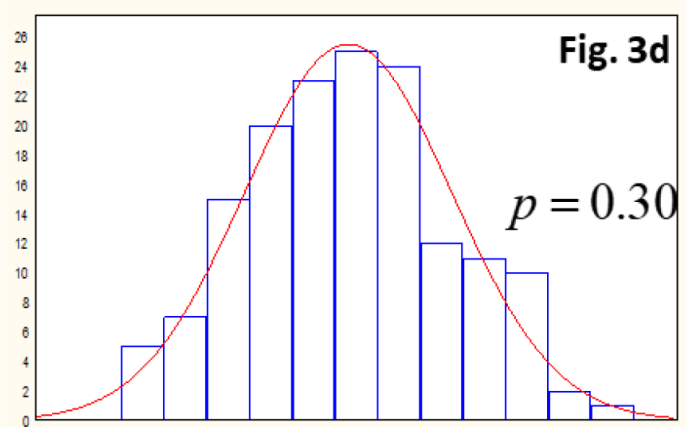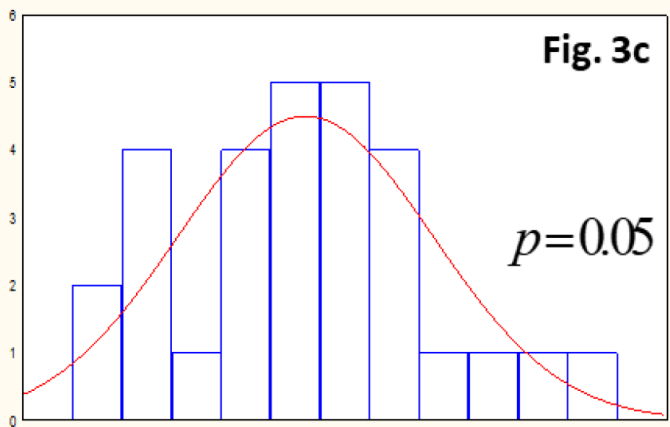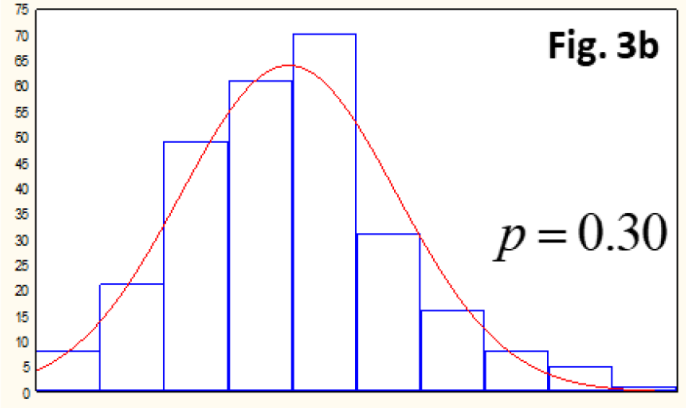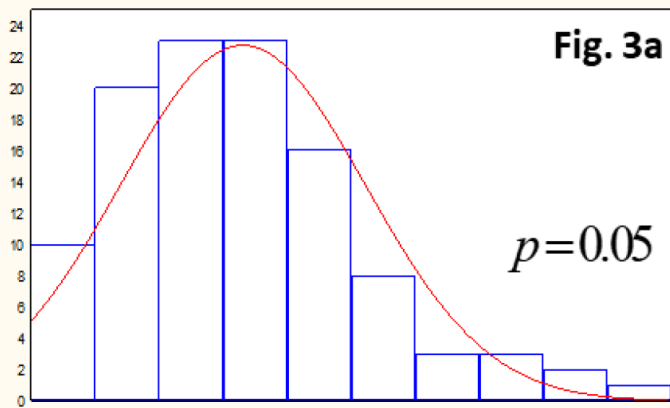
Fig. 3. Histograms of the estimated optimum and decorrelated constrained multistage linear phenotypic selection index (OCMLPSI and DCMLPSI, respectively) values at Stage 2 for simulated dataset in Cycle 2, and the fixed total proportion retained (p) = $q_1 q_2$ = 0.05 and 0.30 when the number of genotypes was 110 (Fig. 3a) and 270 (Fig. 3b) for OCMLPSI, and 30 (Fig. 3c) and 155 (Fig. 3d) for DCMLPSI.

Selection Intensities

The selection intensities (k_1 and k_2) of both indices had three main parts: the proportions retained (q_1 and q_2), the truncation points (u_1 and u_2) and the height of the ordinate of the normal curve [$z(u_1) = e^{-0.5u_1^2}/\sqrt{2\pi}$ and $z(u_2) = e^{-0.5u_2^2}/\sqrt{2\pi}$] (Fig. 1). We obtained the k_1 and k_2 values for OCMLPSI with Eq. [A6] (Appendix B) method and with the Xu and Muir (1992) method for DCMLPSI. Both approaches were associated with the maximum total selection response ($\hat{R}_t = k_1\hat{\sigma}_{I_1} + k_2\hat{\sigma}_{I_2}$) value, and the values of k_1 and k_2 were affected by the method used to obtain their values at Stages 1 and 2. When the p values changed from 0.05 to 0.30, the u_1 and u_2 values decreased, the q_1 and q_2 values increased and the k_1 and k_2 values decreased in both indices, as we would expect.

Equation [A6] (Appendix B) to obtain the OCMLPSI selection intensities in a two-stage context was proposed by Cerón-Rojas et al. (2019). These authors compared their results with the results of Saxton (1983), who used a numerical integration method to obtain truncation points, proportion retained, and selection intensities in a two-stage context. Saxton (1983) applied a two-stage selection scheme in two ways: first, by selecting three

traits and then two traits; and second, by first selecting the last two traits and later the first three traits. Under the first scheme, Saxton (1983) found that the estimated total selection response overestimated the single-stage LPSI response by 3.8%, but under the second, he found that the estimated total selection response overestimated the single-stage LPSI response by only 1.5%. These results were very similar to the results obtained by Cerón-Rojas et al. (2019) when they used real data. This mains that, at least in a two-stage context, Equation [A6] was a good method to obtain the truncation points, proportion retained, and selection intensities.

Number of Restrictions Imposed on the Indices

The OCMLPSI solved the OMLPSI equations subject to the restriction that the covariance between the OCMLPSI and some linear combinations of the genotypes involved be equal to a vector of predetermined proportional gains (or constraints) imposed by the breeder. However, in addition to the latter restriction, the DCMLPSI imposed the restriction that the covariance between DCMLPSI values at different stages be zero. The latter restriction decreased DCMLPSI

efficiency after Stage 1, and as a result, its selection response and expected genetic gain were lower than the OCMLPSI selection response and expected genetic gain for the real and simulated datasets at Stage 2. Xu and Muir (1991, 1992) and Xie and Xu (1997) indicated that the loss of DCMLPSI efficiency after Stage 1 is justified because their method for obtaining the selection intensities and total responses gives the breeder the opportunity to implement an unlimited number of selection stages, which would otherwise be very difficult or impossible to do. At Stage 1, when the additional DCMLPSI restriction was null, the DCMLPSI and OCMLPSI vectors of coefficients were the same, as we would expect. Incidentally, this corroborated that both indices were applications of the SCLPSI to the multistage context.

According to Xu and Muir (1991,1992), the restriction that made the covariance between DCMLPSI values at different stages be zero is similar to the Kempthorne and Nordskog (1959) restriction imposed on the expected genetic gain per trait, which makes some traits not change their mean values while the rest of the trait means remain without restrictions. In effect, the DCMLPSI used a projector matrix (e.g., K_{Di}) to project the OMLPSI vector of coefficients (δ_i) into a space smaller than the original space of δ_i, whereas Kempthorne and Nordskog (1959) used a projector matrix to project the single-stage LPSI vector of coefficients into a space smaller than the original space of the LPSI vector of coefficients. The reduction of the space into which the Kempthorne and Nordskog (1959) matrix projects the LPSI vector of coefficients is equal to the number of zeros that appears in the expected genetic gain per trait, and the selection response and accuracy decrease as the number of restrictions increases (Cerón-Rojas and Crossa, 2018, Chapter 3). Nevertheless, it is not clear whether under the Xu and Muir (1992) restrictions the expected genetic gain per trait, the selection response, and the accuracy decrease as the number of stages increases. If this is true, the Xu and Muir (1992) method could not give the breeder the opportunity to implement an unlimited number of stages, because the expected genetic gain per trait, the selection response, and the accuracy will decrease as the number of stages increases and soon would be null.

In the DMLPSI context, Xie et al. (1997) compared the estimated single-stage LPSI selection response with the estimated DMLPSI selection response for two and three stages and found that at Stages 2 and 3, the estimated total DMLPSI selection response explained only 92 and 87%, respectively, of the estimated LPSI selection response. That is, at Stage 3, the estimated total DMLPSI selection response was lower (5%) than at Stage 2.

Another Way of Writing the OCMLPSI and DCMLPSI Vectors of Coefficients

We wrote the OCMLPSI and DCMLPSI vectors of coefficients ($\beta_i = K_{Oi}\delta_i$ and $b_i = K_{Oi}\delta_i$, respectively) as

a projection of the OMLPSI vector of coefficients ($\delta_i = Q_{ii}^{-1}A_iw$) into a space that is perpendicular to the space generated by the columns of matrix $M_i(V_i)$ made by the projector matrices K_{Oi} and K_{Di} which are idempotent ($K_{Oi} = K_{Oi}^2$ and $K_{Di} = K_{Di}^2$). This is the simplest way of writing the OCMLPSI and DCMLPSI vectors of coefficients. However, there is another way of writing the OCMLPSI and DCMLPSI vectors of coefficients based on the Tallis (1985) approach.

The Tallis (1985) approach requires a proportionality constant which, according to Itoh and Yamada (1987), represents the regression coefficient of the net genetic merit ($H = w'g$) on $Q_{ii}^{-1}V_i(V_i'P^{-1}V_i)^{-1}d_0$, where d_0 is the DCMLPSI vector of predetermined restrictions. There are some problems associated with the proportionality constant. For example, if the proportionality constant is positive, it is appropriate for the DCMLPSI (OCMLPSI) vector of coefficients, and there is no problem; however, if the proportionality constant is negative, the indices will move the population means in the opposite direction to the predetermined desired direction.

Another Constrained Multistage Index

Xie and Xu (1997) developed a constrained multistage selection index as an extension of the DMLPSI developed by Xu and Muir (1992) based on the Tallis (1962) index. Using the Akbar et al. (1984) real data, we found that the Xie and Xu (1997) index was not optimum. The average of the estimated Xie and Xu (1997) selection response for p = 0.05, 0.10, 0.20, and 0.30 explained only 68.55% of the one-stage SCLPSI, whereas the average of the estimated total OCMLPSI and DCMLPSI index selection responses explained 97.60 and 90%, respectively, of the estimated SCLPSI. Similarly, the estimated total Xie and Xu (1997) expected genetic gain values per trait explained only 10.17% of the vector $d' = [3 \quad -1]$ values for both stages. These results indicated that, in effect, the Xie and Xu (1997) index is not optimum and breeders should not use it.

Cerón-Rojas and Crossa (2018) applied the OCMLPSI to the Hicks et al. (1998) dataset, but they used the Young (1964) method to obtain the selection intensities for two stages; thus, their results were approximations because the Young (1964) method overestimated the selection intensities (see Cerón-Rojas et al., 2019, for details).

The Multivariate Normality Assumption of Both Indices

The multivariate normality assumption of the estimated OCMLPSI (DCMLPSI) values was the basis for developing the OCMLPSI (DCMLPSI) theory. Under this assumption, the total OCMLPSI selection response and expected genetic gain per trait for two or more stages, is the sum of each response and expected genetic gain per trait obtained at each stage. We corroborated the normality assumption

at Stage 2 using histograms and normality tests. When at Stage 1 the number of genotypes was 500 and the total proportion retained was 5 or 30%, at Stage 2 the estimated OCMLPSI values approach the normal distribution in a similar manner as the estimated DCMLPSI values. These results indicate that the correlations between the estimated OCMLPSI values do not affect the normality distribution of the estimated OCMLPSI values, at least for the simulated dataset. This means that when the size of the population at Stage 1 is high (e.g., 500 or more), the correlations between the estimated OCMLPSI values cannot affect the normality distribution of the estimated OCMLPSI values in a two-stage context.

CONCLUSIONS

We described the OCMLPSI and DCMLPSI theory and evaluated it in a two-stage context. Based on the estimated total selection response and the total expected genetic gain per trait of each index, we determined their efficiency using a real and a simulated dataset. We found that the OCMLPSI is the most efficient index for predicting the net genetic merit, and its accuracy for predicting the selection response and estimating the expected genetic gain per trait was higher than DCMLPSI accuracy for predicting the selection response and estimating the expected genetic gain per trait. Thus, breeders should use the OCMLPSI when making a selection, not the DCMLPSI.

Conflict of Interest

The authors declare that there is no conflict of interest.

Author Contribution Statement

J.J. Cerón-Rojas developed the statistical methods and the theory of both indices and wrote the manuscript. F.H. Toledo developed an R Package to estimate all the parameters needed to perform the data analysis. J. Crossa reviewed and corrected the manuscript. All the authors read and approved the final manuscript.

References

Akbar, M.K., C.Y. Lin, N.R. Gyles, J.S. Gavora, and C.J. Brown. 1984. Some aspects of selection indices with constraints. Poult. Sci. 63:1899–1905. doi:10.3382/ps.0631899

Ceron-Rojas, J.J., J. Crossa, V.N. Arief, K. Basford, J. Rutkoski, D. Jarquín, et al. 2015. A genomic selection index applied to simulated and real data. G3: Genes, Genomes, Genet. 5:2155–2164. doi:10.1534/g3.115.019869

Cerón-Rojas, J.J., and J. Crossa. 2018. Linear selection indices in modern plant breeding. Springer, Cham, Switzerland. doi:10.1007/978-3-319-91223-3

Cerón-Rojas, J.J., F.H. Toledo, and J. Crossa. 2019. The relative efficiency of two multi-stage linear phenotypic selection indices to predict the net genetic merit. Crop Sci. 59:1037–1051. doi:10.2135/cropsci2018.11.0678

Cochran, W.G. 1951. Improvement by means of selection In: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. Univ. California Press, Berkeley, CA. p. 449–470. https://projecteuclid.org/euclid.bsmsp/1200500247 (accessed 7 June 2019).

Cunningham, E.P. 1975. Multi-stage index selection. Theor. Appl. Genet. 46:55–61. doi:10.1007/BF00264755

Harville, D.A. 1975. Index selection with proportionality constraints. Biometrics 31:223–225. doi:10.2307/2529722

Hazel, L.N. 1943. The genetic basis for constructing selection indexes. Genetics 8:476–490.

Hicks, C., W.M. Muir, and D.A. Stick. 1998. Selection index updating for maximizing rate of annual genetic gain in laying hens. Poult. Sci. 77:1–7. doi:10.1093/ps/77.1.1

Itoh, Y., and Y. Yamada. 1987. Comparisons of selection indices achieving predetermined proportional gains. Genet. Sel. Evol. 19:69–82. doi:10.1186/1297-9686-19-1-69

Kempthorne, O., and A.W. Nordskog. 1959. Restricted selection indices. Biometrics 15:10–19. doi:10.2307/2527598

Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756.

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Assoc., Sunderland, MA.

Mallard, J. 1972. The theory and computation of selection indices with constraints: A critical synthesis. Biometrics 28:713–735. doi:10.2307/2528758

Mardia, K.V. 1980. Tests of univariate and multivariate normality In: P.R. Krishnaiah, editor, Handbook of Statistics 1: Analysis of variance. North-Holland Publ., Amsterdam. p. 279–320.

Mohd-Razali, N., and Y. Bee-Wah. 2011. Power comparisons of Shapiro–Wilk, Kolmogorov–Smirnov, Lilliefors and Anderson–Darling tests. J. Stat. Modeling Analytics 2:21–33.

Moreau, L., A. Charcosset, F. Hospital, and A. Gallais. 1998. Marker-assisted selection efficiency in populations of finite size. Genetics 148:1353–1365.

Namkoong, G. 1970. Optimum allocation of selection intensity in two stages of truncation selection. Biometrics 26:465–476. doi:10.2307/2529102

Podlich, D.W., and M. Cooper. 1998. QU-GENE: A simulation platform for quantitative analysis of genetic models. Bioinformatics 14:632–653. doi:10.1093/bioinformatics/14.7.632

Rani Das, K., and A.H.M. Rahmatullah Imon. 2016. A brief review of tests for normality. Am. J. Theor. Appl. Stat. 5:5–12. doi:10.11648/j.ajtas.20160501.12

Saxton, A.M. 1983. A comparison of exact and sequential methods in multi-stage index selection. Theor. Appl. Genet. 66:23–28. doi:10.1007/BF00281843

Shapiro, S.S., and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). Biometrika 52:591–611. doi:10.2307/2333709

Smith, F.H. 1936. A discriminant function for plant selection. Ann. Eugen. 7:240–250. doi:10.1111/j.1469-1809.1936.tb02143.x

Springer, M.D. 1979. The algebra of random variables. John Wiley & Sons, New York.

Tallis, G.M. 1962. A selection index for optimum genotype. Bio-
metrics 18:120–122. doi:10.2307/2527716

Tallis, G.M. 1985. Constrained selection. Jpn. J. Genet. 60:151–
155. doi:10.1266/jjg.60.151

Xie, C., and S. Xu. 1997. Restricted multistage selection indices.
Genet. Sel. Evol. 29:193–203. doi:10.1186/1297-9686-29-2-
193

Xie, C., S. Xu, and J.A. Mosjidis. 1997. Multistage selection indi-
ces for maximum genetic gain and economic efficiency in red
clover. Euphytica 98:75–82. doi:10.1023/A:1003074814916

Xu, S., and W.M. Muir. 1991. Multistage selection for genetic gain
by orthogonal transformation. Genetics 129:963–974.

Xu, S., and W.M. Muir. 1992. Selection index updating. Theor.
Appl. Genet. 83:451–458. doi:10.1007/BF00226533

Young, S.S.Y. 1964. Multi-stage selection for genetic gain. Hered-
ity 19:131–145. doi:10.1038/hdy.1964.11

APPENDIX A

The Phenotypic Model to Estimate the Variance Components

In this work, we estimated matrices $\mathbf{P}$ and $\mathbf{G}$ (Eq. [1a]) using restricted maximum likelihood (REML) because this estimation method does not require a specific design or balanced data and can be used to estimate genetic and residual variance and covariance in any arbitrary pedigree of individuals. In addition, the expectation and maximization algorithm allows computing the REML for the variance components (Lynch and Walsh, 1998, Chapter 27; Cerón-Rojas and Crossa, 2018, Chapter 2).

Let

$$\mathbf{y}_q = \mathbf{1}\mu_q + \mathbf{Z}\mathbf{g}_q + \mathbf{e}_q$$

be the phenotypic model where $\mathbf{y}_q$ is a $g \times 1$ (g = the number of genotypes in the population) vector of phenotypic averages, which has multivariate normal distribution (NMV) with mean $\mathbf{1}\mu_q$ and covariance matrix $\mathbf{V}_q$; $\mathbf{1}$ is a $g \times 1$ vector of ones, μ_q is the mean of the qth trait, $\mathbf{Z}$ is an identity matrix $g \times g$; $\mathbf{g}_q \sim \text{NMV}(\mathbf{0}, \mathbf{A}\sigma_{g_q}^2)$ is a vector of true breeding values, and $\mathbf{e}_q \sim \text{NMV}(\mathbf{0}, \mathbf{I}\sigma_{e_q}^2)$ is a $g \times 1$ vector of residuals. Matrix $\mathbf{A}$ denotes the numerical relationship matrix between individuals (Lynch and Walsh, 1998), and $\mathbf{V}_q = \mathbf{A}\sigma_{g_q}^2 + \mathbf{I}\sigma_{e_q}^2$. We estimated $\sigma_{g_q}^2$ and $\sigma_{e_q}^2$ in the absence of dominance and epistatic effects.

Estimating Matrices G and P using the Expectation and Maximization Algorithm

The expectation and maximization algorithm allows computing the REML for the variance components $\sigma_{g_q}^2$ and $\sigma_{e_q}^2$ by iterating the following equations:

$$\sigma_{g_q}^{2(n+1)} = \sigma_{g_q}^{2(n)} + \frac{\left[\sigma_{g_q}^{2(n)}\right]^2}{g}$$
$$\times \left\{ \mathbf{y}_q' \left[\mathbf{T}^{(n)}\mathbf{A}\mathbf{T}^{(n)}\right] \mathbf{y}_q - \text{tr}\left[\mathbf{T}^{(n)}\mathbf{A}\right] \right\} \quad \text{[A1]}$$

and

$$\sigma_{e_q}^{2(n+1)} = \sigma_{e_q}^{2(n)} + \frac{\left[\sigma_{e_q}^{2(n)}\right]^2}{g}$$
$$\times \left\{ \mathbf{y}_q' \left[\mathbf{T}^{(n)}\mathbf{T}^{(n)}\right] \mathbf{y}_q - \text{tr}\left[\mathbf{T}^{(n)}\right] \right\} \quad \text{[A2]}$$

where, after n iterations, $\sigma_{g_q}^{2(n+1)}$ and $\sigma_{e_q}^{2(n+1)}$ are the estimated variance components of $\sigma_{g_q}^2$ and $\sigma_{e_q}^2$, respectively. In Eq. [A1] and [A2], tr(.) denotes the trace of the matrices within parentheses; $\mathbf{T} = \mathbf{V}_q^{-1} - \mathbf{V}_q^{-1}\mathbf{1}(\mathbf{1}'\mathbf{V}_q^{-1}\mathbf{1})\mathbf{1}'\mathbf{V}_q^{-1}$ and $\mathbf{V}_q^{-1}$ is the inverse of matrix $\mathbf{V}_q = \mathbf{A}\sigma_{g_q}^2 + \mathbf{I}\sigma_{e_q}^2$. In $\mathbf{T}^{(n)}$, $\mathbf{V}_q^{-1(n)}$ is the inverse of matrix $\mathbf{V}_q^{(n)} = \sigma_{g_q}^{2(n)} + \mathbf{I}\sigma_{e_q}^{2(n)}$.

The additive genetic and residual covariances between the observations of the qth and ith traits, $\mathbf{y}_q$ and $\mathbf{y}_i$ ($\sigma_{g_{q,i}}$ and $\sigma_{e_{q,i}}$, q, $i = 1, 2, \ldots, t$), can be estimated with REML by adapting Eq. [A1] and [A2] as follows. The variance of the sum of $\mathbf{y}_q$ and $\mathbf{y}_i$ can be written as $\text{Var}(\mathbf{y}_i + \mathbf{y}_q) = \mathbf{V}_i + \mathbf{V}_q + 2\mathbf{C}_{iq}$, where $\mathbf{V}_i = \mathbf{A}\sigma_{g_i}^2 + \mathbf{I}\sigma_{e_i}^2$ is the variance of $\mathbf{y}_i$ and $\mathbf{V}_q = \mathbf{A}\sigma_{g_q}^2 + \mathbf{I}\sigma_{e_q}^2$ is the variance of $\mathbf{y}_q$; in addition, $2\mathbf{C}_{iq} = 2\mathbf{A}\sigma_{giq} + 2\mathbf{I}\sigma_{eiq} = 2\text{Cov}(\mathbf{y}_i, \mathbf{y}_q)$ is the covariance of $\mathbf{y}_q$ and $\mathbf{y}_i$, and σ_{giq} and σ_{eiq} are the additive and residual covariances, respectively, associated with the covariance of $\mathbf{y}_q$ and $\mathbf{y}_i$. Thus, one way of estimating σ_{giq} and σ_{eiq} is using the following equation:

$$0.5\text{Var}\left(\mathbf{y}_i + \mathbf{y}_q\right) - 0.5\text{Var}\left(\mathbf{y}_i\right) - 0.5\text{Var}\left(\mathbf{y}_q\right) \quad \text{[A3]}$$

for which Eq. [A1] and [A2] can be used.

Estimating the OCMLPSI and DCMLPSI Selection Response and Expected Genetic Gain per Trait at Stage *i*

By Eq. [A1] to [A3], the estimates of matrices $\mathbf{P}$ and $\mathbf{G}$ can be denoted as $\hat{\mathbf{P}}$ and $\hat{\mathbf{G}}$, and the estimated OCMLPSI and DCMLPSI selection responses (Eq. [8a] and [8b]) at Stage i are

$$\hat{R}_{Oi} = k_{Oi}\sqrt{\hat{\boldsymbol{\beta}}_i'\hat{\mathbf{Q}}_{ii}\hat{\boldsymbol{\beta}}_i} \quad \text{[A4a]}$$

and

$$\hat{R}_{Di} = k_{Di}\sqrt{\hat{\mathbf{b}}_i'\hat{\mathbf{Q}}_{ii}\hat{\mathbf{b}}_i} \quad \text{[A4b]}$$

whereas the estimated OCMLPSI and DCMLPSI expected genetic gains per trait (Eq. [9a] and [9b]) at Stage i are

$$\hat{\mathbf{E}}_{Oi} = k_{Oi}\frac{\hat{\mathbf{A}}_i'\hat{\boldsymbol{\beta}}_i}{\sqrt{\hat{\boldsymbol{\beta}}_i'\hat{\mathbf{Q}}_{ii}\hat{\boldsymbol{\beta}}_i}} \quad \text{[A5a]}$$

and

$$\hat{\mathbf{E}}_{Di} = k_{Di}\frac{\hat{\mathbf{A}}_i'\hat{\mathbf{b}}_i}{\sqrt{\hat{\mathbf{b}}_i'\hat{\mathbf{Q}}_{ii}\hat{\mathbf{b}}_i}} \quad [\text{A5b}]$$

respectively.

APPENDIX B
The OCMLPSI Selection Intensity for Two Stages

We describe a method to obtain the OCMLPSI selection intensity for fixed total proportion $p = q_1 q_2$, where q_1 and q_2 are the proportions of individuals selected at Stage 1 and 2, respectively. This is because Cerón-Rojas et al. (2019) found that the Cochran (1951) and Young (1964) method overestimates the selection intensity. We obtained the DCMLPSI selection intensity according to the Xu and Muir (1992) method.

Let $I_1 = \boldsymbol{\beta}'_1\mathbf{x}_1$ and $I_2 = \boldsymbol{\beta}'_2\mathbf{y}$ be the OCMLPSI at Stages 1 and 2, respectively, and assume that the indices have bivariate normal distribution. Let I_1 and I_2 be transformed as $u_1 = \left(I_1 - \mu_{I_1}\right)/\sigma_{I_1}$ and $u_2 = \left(I_2 - \mu_{I_2}\right)/\sigma_{I_2}$ with mean zero and variance 1.0, where μ_{I_1} and μ_{I_2} are the means, whereas σ_{I_1} and σ_{I_2} are the standard deviations of I_1 and I_2, respectively. The selected population has bivariate left truncated normal distribution with probability density function $h(u_1, u_2) = f(u_1, u_2)/p$, where $p = q_1 q_2$,

$$f(u_1, u_2) = \frac{1}{2\pi\sqrt{1-\rho_{12}^2}}\exp\left[-\frac{1}{2(1-\rho_{12}^2)}\left(u_1^2 + u_2^2 - 2\rho_{12}u_1u_1\right)\right]$$

and ρ_{12} is the correlation between u_1 and u_2 (Young, 1964; Cerón-Rojas and Crossa 2018, Chapters 2 and 9).

Consider the transformations (Springer, 1979) $v_1 = u_1$ and $v_2 = (u_2 - \rho_{12}u_1)/\sqrt{1-\rho_{12}^2}$, with Jacobian $\mathbf{j}$, where

$$\mathbf{j}^{-1} = \begin{vmatrix} \dfrac{\partial v_1}{\partial u_1} & \dfrac{\partial v_1}{\partial u_2} \\ \dfrac{\partial v_2}{\partial u_1} & \dfrac{\partial v_2}{\partial u_2} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ \dfrac{-\rho_{12}}{\sqrt{1-\rho_{12}^2}} & \dfrac{1}{\sqrt{1-\rho_{12}^2}} \end{vmatrix} = \dfrac{1}{\sqrt{1-\rho_{12}^2}}$$

where $|\text{o}|$ denotes the determinant function and ∂ denotes the partial derivatives of v_1 and v_2 with respect to u_1 and u_2. Thus,

$$v_1^2 + v_2^2 = \frac{u_1^2 + u_2^2 - 2\rho_{12}u_1u_2}{(1-\rho_{12}^2)}$$

and

$$g(v_1, v_2) = |\mathbf{j}|f(u_1, u_2) = \left(\frac{1}{\sqrt{2\pi}}e^{-0.5v_1^2}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-0.5v_2^2}\right)$$

The transformations indicate that variables v_1 and v_2 are independent, each with a standard normal distribution.

Variables v_1 and v_2 are associated with the truncation points as $u_1 = v_1$ and $u_2 = v_2\sqrt{1-\rho_{12}^2} + \rho_{12}v_1$. This means that u_1 and u_2 values should be obtained in two steps. First, we obtained the values of v_1 and v_2 from two independent standard normal distributions; then, we obtained the values of $v_1 = u_1$, $u_2 = v_2\sqrt{1-\rho_{12}^2} + \rho_{12}v_1$, q_1, and q_2; finally, we obtained the OCMLPSI selection intensity at Stages 1 and 2 as

$$k_1 = \frac{z(u_1)}{q_1} \quad \text{and} \quad k_2 = \frac{z(u_2)}{q_2} \quad [\text{A6}]$$

respectively, where $z(u_1) = e^{-0.5u_1^2}/\sqrt{2\pi}$ and $z(u_2) = e^{-0.5u_2^2}/\sqrt{2\pi}$ were the height of the ordinate of the normal curve at the lowest values of u_1 and u_2 retained, whereas q_1 and q_2 are the proportions of the population of animals or plants selected at each stage (Fig. 2). Equation [A6] values should be associated with the maximum $R_t = R_1 + R_2$ value (Fig. 2), where $R_1 = k_1\sigma_{I_1}$ and $R_2 = k_2\sigma_{I_2}$ are the selection responses, whereas σ_{I_1} and σ_{I_2} are the standard deviations of the variance of I_1 and I_2 at Stages 1 and 2, respectively.