Published June 29, 2016

RESEARCH

A Predetermined Proportional Gains Eigen Selection Index Method

J. Jesus Cerón-Rojas, José Crossa,* Fernando H. Toledo, and Jaime Sahagún-Castellanos

ABSTRACT

The most general linear phenotypic selection index (PSI) is the predetermined proportional gains phenotypic selection index (PPG-PSI) that allows imposing restrictions on the trait expected genetic gain values to make some traits change their mean values based on a predetermined level, while the rest of the traits remain without restrictions. However, due to the increasing number of restricted traits: (i) PPG-PSI accuracy decreases; (ii) the proportional constant associated with this index can be negative, in which case, its results have no meaning in practice; and (iii) the PPG-PSI can shift the population means in the opposite direction to the predetermined desired direction. Based on the eigen selection index method (ESIM), we propose a PPG-ESIM that does not require a proportional constant, and due to the properties associated with eigen analysis, it is possible to use the theory of similar matrices to change the direction of the eigenvector values without affecting PPG-ESIM accuracy, which helps to eliminate the problem indicated in the third point above, associated with the standard PPG-PSI. The PPG-ESIM uses the first eigenvector as its vector of coefficients, and the first eigenvalue in the selection response. Two simulated and one real data set, each with four traits, were used to validate PPG-ESIM efficiency vs. PPG-PSI efficiency; the simulated and real results indicated that PPG-ESIM efficiency was higher than PPG-PSI efficiency. We concluded that PPG-ESIM is an efficient selection index that can be used in any selection program as a good alternative to PPG-PSI.

J.J. Cerón-Rojas and J. Sahagún-Castellanos, Instituto de Horticultura, Dep. de Fitotecnia, Univ. Autónoma Chapingo, Chapingo, México, C.P. 56230; J. Crossa and F.H. Toledo, Biometrics and Statistics Unit (BSU), International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, 06600, México D.F., México. Received 26 Nov. 2015. Accepted 25 May 2016. Assigned to Associate Editor Natalia de Leon. *Corresponding author (j.crossa@cgiar.org).

Abbreviations: ESIM, eigen selection index method; *H*, net genetic merit; *I*, linear index; PPG-ESIM, predetermined proportional gains eigen selection index method; PPG-PSI, predetermined proportional gains phenotypic selection index; PSI, phenotypic selection indices; RESIM, restricted eigen selection index; RPSI, restricted phenotypic selection index; SPSI, Smith phenotypic selection index.

The most general linear phenotypic selection index (PSI) associated with the standard Smith (1936) PSI (SPSI) is the optimum predetermined proportional gains (PPG) PSI (PPG-PSI), originally developed by Mallard (1972) as an extension of the Kempthorne and Nordskog (1959) restricted PSI (RPSI). The PPG-PSI allows imposing restrictions on the trait expected genetic gains to make some traits change their mean values based on a predetermined level, while the rest of the traits change their values without restriction.

A PSI is a linear combination of several observable, optimally weighted trait values; its aims are to: (i) predict the unobservable net genetic merit values of the candidates for selection; (ii) maximize the selection response and the expected genetic gains for each trait; and (iii) provide the breeder with an objective rule for evaluating and selecting several traits simultaneously (Baker, 1974). Smith (1936) developed the basic theory of the PSI under two assumptions: first, the net genetic merit of the candidates for selection is a linear combination of the additive genetic values of several traits; and second, the PSI and the net genetic merit have a jointly normal

Published in Crop Sci. 56:2436–2447 (2016).
doi: 10.2135/cropsci2015.11.0718

distribution. The main problem of the SPSI is that when used to select individuals, the expected genetic gain values of the individual traits can change in a positive or negative direction without any control. Because of this, Kempthorne and Nordskog (1959) developed the basic theory of the RPSI that allows imposing restrictions equal to zero on the expected genetic gains of some traits, while other traits increase (or decrease) their expected genetic gains without imposing any restrictions. Mallard (1972) extended the original RPSI theory to the case of a PSI that allows imposing optimal predetermined level restrictions on the trait expected genetic gains. The main objective of the Smith (1936) PSI was to maximize the selection response, while the main objective of the RPSI and the PPG-PSI is to optimize, under some restrictions, the expected genetic gains per selection cycle for each trait.

The PPG-PSI is more useful than any other PSI because its expected genetic gain values can change according to the breeder's interest. For example, the expected genetic gain value for one observed trait can be negative in the SPSI, while the breeder's interest may lie in a positive expected genetic gain value, or vice versa. In this case, the PPG-PSI can impose positive or negative restrictions on the expected genetic gain values. By this reasoning, the PPG-PSI should be the basic tool for selecting individuals as parents of the next generation.

Other PPG-PSIs were proposed by Harville (1975) and Tallis (1985). Itoh and Yamada (1987) showed that the Mallard (1972) index is equal to the Tallis (1985) index and that, except for a proportional constant, the Tallis (1985) index is equal to the Harville (1975) index. Thus, according to the results of Itoh and Yamada (1987), in reality there is only one PPG-PSI. From a theoretical point of view, it is easier to work with the Mallard (1972) PPG-PSI than the Harville (1975) or the Tallis (1985) PPG-PSI. That is, it is easier to obtain the sampling statistical properties of the coefficients of the Mallard (1972) PPG-PSI than those of the Harville (1975) or the Tallis (1985) PPG-PSI (Cerón-Rojas et al., 2016).

In the canonical correlations context, a phenotypic unrestricted and restricted eigen selection index method (ESIM and RESIM, respectively), which uses the elements of the first eigenvector as the PSI coefficient and the first eigenvalue in the selection response, was developed by Cerón-Rojas et al. (2008a). The sampling statistical properties of ESIM are known and Cerón-Rojas et al. (2016) have shown that they can be used to find the sampling statistical properties of the coefficients of the SPSI, the RPSI and the PPG-PSI.

When the trait's phenotypic and genotypic variances and covariances are known, the PPG-PSI has similar statistical properties as the SPSI: (i) the correlation between the net genetic merits of the candidates for selection and the PPG-PSI is maximized; (ii) the mean prediction error is minimized; and (iii) the PPG-PSI is the best linear predictor of the net genetic merit of the candidates for selection and the one with the highest relative efficiency when compared to other selection procedures (Lin, 2005). However, when the phenotypic and genotypic variances and covariances are estimated, the PPG-PSI selection response and expected genetic gain will be optimal only if the estimators of the index coefficients are unbiased and have minimal variance (Hayes and Hill, 1980; Cerón-Rojas et al., 2016). An additional disadvantage of the PPG-PSI is that it is difficult to assign economic weights to the traits of interest.

Authors such as Vandepitte and Hazel (1977) and Smith (1983) have used computer simulation to estimate economic weights; however, their results are limited to some traits of pigs. Using economic theory, Melton et al. (1979) proposed a profit function to estimate the economic weight; however, Goddard (1983) found inconsistencies in the method of Melton et al. (1979). Magnussen (1990) estimated the economic weights as functions of the eigenvalues and eigenvectors of a quadratic form of the additive genetic and phenotypic covariance matrices, but he did not show that his estimated economic weights maximize the correlation between the SPSI and the net genetic merit.

Itoh and Yamada (1987) pointed out several additional problems associated with the PPG-PSI: due to the restrictions imposed on the PPG-PSI expected genetic gains, there is a loss of accuracy in the PPG-PSI, that is, the correlation between the PPG-PSI and the net genetic merit tends to decrease as the number of restricted traits increases; the proportional constant associated with the PPG-PSI can be negative, in which case PPG-PSI results have no meaning in practice; and the PPG-PSI may cause the population means to shift in the opposite direction to the predetermined desired direction; this may happen due to the opposite directions between the economic values and the predetermined desired direction.

The aims of the present study are to: (i) propose a PPG eigen selection index method (PPG-ESIM) as a generalization of the ESIM (RESIM) of Cerón-Rojas et al. (2008a); (ii) show that the PPG-ESIM does not require a proportional constant; (iii) show that due to the properties associated with eigen analysis, it is possible to use the theory of similar matrices (Harville, 1997) to change the direction of the eigenvector values without affecting the correlation between the PPG-ESIM and the net genetic merit, which helps to eliminate the problem of PPG-PSI in the third point indicated by Itoh and Yamada in the previous paragraph; (iv) estimate the economic weight in the canonical correlation context and show that the estimated economic weights maximize the correlation between the PPG-ESIM and the net genetic merit; and (v) compare the efficiency of PPG-PSI to PPG-ESIM efficiency.

Like ESIM (RESIM), PPG-ESIM uses the elements of the first eigenvector as the vector of coefficients and the first eigenvalue in the selection response. The efficiency of PPG-ESIM is expected to be higher than PPG-PSI efficiency in all selection cycles, regardless of the number of restricted traits.

MATERIALS AND METHODS

Linear Phenotypic Selection Index (PSI) Theory

Objectives of the PSI

The objectives of any linear phenotypic selection index are: first to predict the net genetic merit $H = \mathbf{w}'\mathbf{a}$, where $\mathbf{a}' = [a_1\ a_2 \ldots a_t]$ (t = number of traits) is the vector of true breeding values for an individual and $\mathbf{w}' = [w_1\ w_2 \ldots w_t]$ is the vector of economic weights; and second to select individuals with the highest H values in each selection cycle as parents of the next generation.

Expected Genetic Gain per Selection Cycle for Each Trait and Selection Response

While the objective of the SPSI is to maximize the selection response, the main objective of PPG-PSI is to optimize, under some restrictions, the expected genetic gain per selection cycle for each trait by requiring that its values change to some predetermined value, or in some specified direction (sign). According to Kempthorne and Nordskog (1959), the expected genetic gain per selection cycle for each trait ($\mathbf{E}$) of any linear PSI can be written as

$$\mathbf{E} = k\frac{\mathbf{Cb}}{\sigma_I} \qquad [1a]$$

where k is the standardized selection differential (or selection intensity), $\mathbf{C}$ is the matrix of covariance of true breeding values $\mathbf{a}$; $\mathbf{b}$ and σ_I are the vectors of coefficients and the standard deviation of the variance (σ_I^2) of any linear index $I = \mathbf{b}'\mathbf{p}$, respectively; $\mathbf{p}' = [p_1\ p_2 \ldots p_t]$ is a vector of trait phenotypic values, and t = number of traits in the selection index. In the SPSI context, vector $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Cw}$ optimizes Eq. [1a], where $\mathbf{P}^{-1}$ is the inverse of the phenotypic covariance matrix, $\mathbf{P}$.

Selection Response

For any linear selection index, the equation that predicts the net genetic change due to selection for $H = \mathbf{w}'\mathbf{a}$ from one selection cycle to the next, is the selection response, which can be written as

$$R = k\frac{\sigma_{H,I}}{\sigma_I} = k\sigma_H\rho_{H,I} \qquad [1b]$$

where k and σ_I were defined in Eq. [1a], σ_H is the standard deviation of H, $\sigma_{H,I}$ and $\rho_{H,I}$ are the covariance and the correlation between H and any linear index I, respectively. The second part of Eq. [1b] ($k\sigma_H\rho_{H,I}$) indicates that the genetic change due to selection is proportional to $\rho_{H,I}$ and to k (Kempthorne and Nordskog, 1959). In general, it is assumed that k and σ_H are fixed and then, R will be maximized when $\rho_{H,I}$ is maximized and the final form of Eq. [1b] will depend on the particular linear selection index used to select individuals, e.g., PPG-PSI or PPG-ESIM.

Predetermined Proportional Gains Phenotypic Selection Index (PPG-PSI) Theory

Mallard (1972) extended the Kempthorne and Nordskog (1959) RPSI by considering that if μ_q is the population mean of the qth trait before selection, one objective could be to change μ_q to $\mu_q + d_q$, where d_q is the predetermined proportional gain imposed by the breeder on μ_q in one selection cycle (in Kempthorne and Nordskog [1959], $d_q = 0$, $q = 1, 2, \ldots, r$); the rest of the traits change without any restrictions.

Vector of Coefficients of the PPG-PSI

Let $\mathbf{d}' = [d_1\ d_2 \ldots d_s]$ be the vector of the predetermined proportional gains imposed by the breeder, and

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

a matrix $r(r-1)$ where r is the number of predetermined proportional gains and d_q (for $q = 1, 2, \ldots, r$) is the qth element of vector $\mathbf{d}'$. Matrix $\mathbf{D}'$ is the Mallard (1972) matrix of predetermined proportional gains and can be used to impose predetermined restrictions on the expected genetic gain per selection cycle for each trait (Eq. [1a]). Let $\mathbf{M}' = \mathbf{D}'\Psi'$ be the Mallard (1972) matrix of predetermined restrictions, where $\Psi' = \mathbf{U}'\mathbf{C}$ and $\mathbf{U}'$ is the Kempthorne and Nordskog (1959) matrix of restrictions comprised of values 1 and 0, where '1' indicates that the trait is restricted (i.e., $d_q = 0$) and '0' that the trait has no restrictions. The vector of predetermined restrictions $\mathbf{M}'\mathbf{b} = 0$ is used to maximize the correlation between $I = \mathbf{b}'\mathbf{p}$ and $H = \mathbf{w}'\mathbf{a}$.

Maximizing $\Phi_M = \mathbf{w}'\mathbf{Cb} - 0.5\lambda(\mathbf{b}'\mathbf{Pb} - 1) - \mathbf{v}'\mathbf{M}'\mathbf{b}$ with respect to $\mathbf{b}'$, $\mathbf{v}'$, and 0.5λ, where 0.5λ and $\mathbf{v}' = [v_1\ v_2 \ldots v_{r-1}]$ are Lagrange multipliers, is equivalent to maximizing the correlation between $I = \mathbf{b}'\mathbf{p}$ and $H = \mathbf{w}'\mathbf{a}$ (Bulmer, 1980). The vector of coefficients of the Mallard (1972) PPG-PSI is

$$\mathbf{b}_M = \mathbf{Kb} \qquad [2a]$$

where $\mathbf{K} = [\mathbf{I}_t - \mathbf{Q}]$, $\mathbf{Q} = \mathbf{P}^{-1}\Psi\mathbf{D}(\mathbf{D}'\Psi'\mathbf{P}^{-1}\Psi\mathbf{D})^{-1}\mathbf{D}'\Psi'$, $\mathbf{I}_t$ is an identity matrix $t \times t$, and $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Cw}$ is the vector of coefficients of the SPSI. When $\mathbf{D}' = \mathbf{U}'$, $\mathbf{b}_M = \mathbf{b}_{KN}$ (the vector of the RPSI), and when $\mathbf{U}'$ is a null matrix, $\mathbf{b}_M = \mathbf{b}$. That is, the Mallard (1972) index is more general than the Kempthorne and Nordskog (1959) RPSI and is an optimum PPG-PSI.

According to Itoh and Yamada (1987), an additional form of writing Eq. [2a] is as

$$\mathbf{b}_M = \mathbf{b}_{KN} + \theta\delta \qquad [2b]$$

where $\mathbf{b}_{KN} = [\mathbf{I} - \mathbf{P}^{-1}\Psi(\Psi'\mathbf{P}^{-1}\Psi)^{-1}\Psi']\mathbf{b}$ is the vector of coefficients of the Kempthorne and Nordskog (1959) RPSI, $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Cw}$, $\delta = \mathbf{P}^{-1}\Psi(\Psi'\mathbf{P}^{-1}\Psi)^{-1}\mathbf{d}$, and

$$\theta = \frac{\mathbf{b}'\Psi(\Psi'\mathbf{P}^{-1}\Psi)^{-1}\mathbf{d}}{\mathbf{d}'(\Psi'\mathbf{P}^{-1}\Psi)^{-1}\mathbf{d}} \qquad [3]$$

is a proportional constant associated with the PPG-PSI coefficient. When $\theta = 0$, $\mathbf{b}_M = \mathbf{b}_{KN}$, and if $\mathbf{U}'$ is the null matrix,

$\mathbf{b}_M = \mathbf{b}$. Equation [2b] is the Tallis (1985) PPG-PSI vector of coefficients; that is, the Mallard (1972) and Tallis (1985) indices are the same.

Maximized PPG-PSI Expected Genetic Gains and Selection Response

The PPG-PSI expected genetic gains per selection cycle for each trait can be written as

$$\mathbf{E}_{\text{PSI}} = k\frac{\mathbf{Cb}_M}{\sqrt{\mathbf{b}'_M\mathbf{Pb}_M}} \quad [4a]$$

Equations [2a] and [2b] maximize the selection response, which can be written as

$$R_{\text{PSI}} = k\sqrt{\mathbf{b}'_M\mathbf{Pb}_M} \quad [4b]$$

The parameters of Eq. [4a] and [4b] were previously defined.

Proposed Predetermined Proportional Gains Eigen Selection Index (PPG-ESIM)

PPG-ESIM Vector of Coefficients

To obtain the PPG-ESIM vector of coefficients, we added to equation $\Phi_M = \mathbf{w}'\mathbf{Cb} - 0.5\lambda(\mathbf{b}'\mathbf{Pb} - 1) - \mathbf{v}'\mathbf{M}'\mathbf{b}$ the restriction $\mathbf{w}'\mathbf{Cw} = 1$, and we maximized equation $\Phi = \mathbf{w}'\mathbf{Cb} - 0.5\lambda(\mathbf{b}'\mathbf{Pb} - 1) - 0.5\mu(\mathbf{wCw} - 1) - \mathbf{v}'\mathbf{M}'\mathbf{b}$ with respect to $\mathbf{b}$, $\mathbf{w}$, 0.5λ, 0.5μ, and $\mathbf{v}'$, where 0.5λ, 0.5μ, and $\mathbf{v}' = [v_1\ v_2\ \dots\ v_{r-1}]$ were Lagrange multipliers. This process maximized the correlation between $I = \mathbf{b}'\mathbf{p}$ and $H = \mathbf{w}'\mathbf{a}$. The PPG-ESIM vector of coefficients can be obtained from equation

$$(\mathbf{T} - \lambda^2\mathbf{I})\mathbf{b}_{\text{PPG-ESIM}} = \mathbf{0} \quad [5a]$$

where $\mathbf{T} = [\mathbf{I}_t - \mathbf{P}^{-1}\Psi\mathbf{D}(\mathbf{D}'\Psi'\mathbf{P}^{-1}\Psi\mathbf{D})^{-1}\mathbf{D}'\Psi']\mathbf{P}^{-1}\mathbf{C} = \mathbf{KP}^{-1}\mathbf{C}$ (matrix $\mathbf{K}$ was defined in Eq. [2a]), λ^2 is the first eigenvalue (λ is the maximum correlation between $I = \mathbf{b}'\mathbf{p}$ and $H = \mathbf{w}'\mathbf{a}$) and $\mathbf{b}_{\text{PPG-ESIM}}$ is the first eigenvector of matrix $\mathbf{T}$. When $\mathbf{D}' = \mathbf{U}'$, $\mathbf{T} = [\mathbf{I} - \mathbf{P}^{-1}\Psi(\Psi'\mathbf{P}^{-1}\Psi)^{-1}\Psi']\mathbf{P}^{-1}\mathbf{C}$ (RESIM case), and when $\mathbf{U}'$ is a null matrix, $\mathbf{T} = \mathbf{P}^{-1}\mathbf{C}$ (ESIM case), from where Eq. [5a] is more general than RESIM and ESIM, and these last two indices are particular cases of PPG-ESIM. In Eq. [5a], the values of $\mathbf{b}_{\text{PPG-ESIM}}$ and λ^2 can be estimated using singular value decomposition (Cerón-Rojas et al., 2008a).

$\mathbf{b}_{\text{PPG-ESIM}}$ is Independent of the Economic Values and of the θ Values

Note that as $\mathbf{b}_{\text{PPG-ESIM}}$ is the first eigenvector of matrix $\mathbf{T}$, then it is independent of the economic values and of the θ values (Eq. [3]).

PPG-ESIM Coefficient Vector and the Theory of Similar Matrices

Equation [5a] can be written as $\mathbf{TIb}_{\text{PPG-ESIM}} = \lambda^2\mathbf{Ib}_{\text{PPG-ESIM}}$, where $\mathbf{I} = \mathbf{F}^{-1}\mathbf{F}$ is an identity matrix and $\mathbf{F} = diag\{f_1\ f_2\ \dots\ f_t\}$ (t = number of traits) is a diagonal matrix with values equal to any real number, except zero values. Then, an additional form of write Eq. [5a] is as

$$(\mathbf{T}_2 - \lambda^2\mathbf{I})\beta = \mathbf{0} \quad [5b]$$

where $\mathbf{T}_2 = \mathbf{FTF}^{-1}$ and $\beta = \mathbf{Fb}_{\text{PPG-ESIM}}$; $\mathbf{T}$ and $\mathbf{T}_2 = \mathbf{FTF}^{-1}$ are similar matrices and both have the same eigenvalues but different eigenvectors (Harville, 1997). When the $\mathbf{F}$ values are only 1s, vector $\mathbf{b}_{\text{PPG-ESIM}}$ is not affected; when the $\mathbf{F}$ values are only −1s, vector $\mathbf{b}_{\text{PPG-ESIM}}$ will change its direction, and if the $\mathbf{F}$ values are different from 1 and −1, matrix $\mathbf{F}$ will change the proportional values of $\mathbf{b}_{\text{PPG-ESIM}}$. In practice, first we obtained $\mathbf{b}_{\text{PPG-ESIM}}$ from $(\mathbf{T} - \lambda^2\mathbf{I})\mathbf{b}_{\text{PPG-ESIM}} = 0$ and then we multiplied $\mathbf{b}_{\text{PPG-ESIM}}$ by the $\mathbf{F}$ values to obtain $\beta = \mathbf{Fb}_{\text{PPG-ESIM}}$, that is, we did a linear transformation of $\mathbf{b}_{\text{PPG-ESIM}}$. Cerón-Rojas et al. (2006) introduced an alternative procedure for modifying the $\mathbf{b}_{\text{PPG-ESIM}}$ signs that is a particular case of the method proposed in this work based on the theory of similar matrices.

Estimating the PPG-ESIM Vector of Economic Weights

In the context of canonical correlation, Cerón-Rojas et al. (2008a) showed that the vector of economic weights can be written as

$$\mathbf{w}_E = \mathbf{C}^{-1}[\lambda\mathbf{Pb}_{\text{PPG-ESIM}} + \mathbf{Mv}] \quad [6]$$

where $\mathbf{C}$ and λ were defined in Eq. [1a] and [5a], respectively; $\mathbf{M}' = \mathbf{D}'\Psi'$ is the Mallard (1972) matrix of predetermined restrictions (Eq. [2a]), $\Psi' = \mathbf{U}'\mathbf{C}$ was defined in Eq. [2a], and $\mathbf{v} = \lambda^{-1}(\mathbf{M}'\mathbf{P}^{-1}\mathbf{M})^{-1}\mathbf{M}'\mathbf{P}^{-1}\mathbf{Cb}_{\text{PPG-ESIM}}$.

PPG-ESIM Expected Genetic Gains and Selection Response

The PPG-ESIM expected genetic gains per selection cycle for each trait can be written as

$$\mathbf{E}_{\text{ESIM}} = k\frac{\mathbf{C}\beta}{\sqrt{\beta\mathbf{P}\beta}} \quad [7a]$$

and, by Eq. [1b], [5a] and [6], the PPG-ESIM selection response can be written as

$$R_{\text{ESIM}} = k\sigma_{HE}\lambda \quad [7b]$$

where $\sigma_{HE} = \sqrt{\mathbf{w}'_E\mathbf{Cw}_E}$ and λ was defined in Eq. [5a]. Note that Eq. [7a] does not requires economic weights, and when $\sigma_{HE} = 1$, Eq. [7b] can be written as $R_{\text{ESIM}} = k\lambda$ (Cerón-Rojas et al., 2008a); that is, in this last form it does not require economic weights.

Criteria for Comparing PPG-ESIM efficiency vs. PPG-PSI Efficiency

To compare PPG-ESIM efficiency vs. PPG-PSI efficiency, we used the ratio

$$\hat{\pi} = \frac{\hat{\rho}_{H,\text{ESIM}}}{\hat{\rho}_{H,\text{PSI}}} \quad [8]$$

which was proposed by Bulmer (1980) as a criterion for comparing the efficiency of linear PSI. In Eq. [8], $\hat{\rho}_{H,\text{ESIM}} = \hat{\lambda}$ and

$$\hat{\rho}_{H,\text{PSI}} = \frac{\mathbf{w}'\hat{\mathbf{C}}\hat{\mathbf{b}}_M}{\sqrt{\mathbf{w}'\hat{\mathbf{C}}\mathbf{w}}\sqrt{\hat{\mathbf{b}}'_M\hat{\mathbf{P}}\hat{\mathbf{b}}_M}} \quad [9]$$

are the maximized estimated correlations (or accuracy) between H and PPG-ESIM, and between H and PPG-PSI, respectively, where $\hat{\lambda}$ is an estimation of the square root of the first eigenvalue in Eq. [5a], $\hat{\mathbf{C}}$ and $\hat{\mathbf{P}}$ are estimates of matrices $\mathbf{C}$ and $\mathbf{P}$, and $\hat{\mathbf{b}}_M$ is an estimation of Eq. [2a] or [2b]. Using this criterion, if $\hat{\pi} > 1$, PPG-ESIM efficiency will be greater than PPG-PSI efficiency, if $\hat{\pi} = 1$, the efficiency of both selection indices will be equal, and if $\hat{\pi} < 1$, PPG-PSI will be more efficient than PPG-ESIM.

Simulated and Real Data Sets

Simulated Data Set 1

This data set was simulated by Ceron-Rojas et al. (2015) and can be obtained at http://hdl.handle.net/11529/10199. The data were simulated for 8 phenotypic selection cycles (C0 to C7) each with 4 traits (T1, T2, T3 and T4), 500 genotypes and 4 replicates for each genotype. In this paper, we used only cycles C1 to C7.

In all selection cycles, we used a selection intensity of 10% (k = 1.75) to make selections using PPG-PSI and PPG-ESIM. The PPG-PSI economic weights for T1, T2, T3 and T4 were 1, −1, 1, and 1, respectively. In each selection cycle, we estimated the proportional constant associated with the PPG-PSI coefficient (Eq. [3]), the PPG-PSI and PPG-ESIM expected genetic gain per selection cycle for each trait (Eq. [4a] and [7a]), the PPG-PSI and PPG-ESIM selection responses (Eq. [4b] and [7b]), the correlation between the net genetic merit $H = \mathbf{w}'\mathbf{a}$ and PPG-PSI and between $H = \mathbf{w}'\mathbf{a}$ and PPG-ESIM (Eq. [8]), the ratio $\hat{\pi} = \hat{\lambda}/\hat{\rho}_{H,\mathrm{PSI}}$ (Eq. [8]), and finally, we calculated the true correlation between $H = \mathbf{w}'\mathbf{a}$ and the PPG-PSI index and between $H = \mathbf{w}'\mathbf{a}$ and the PPG-ESIM (λ and $\rho_{H,\mathrm{PSI}}$, respectively) index.

Simulated data were generated using QU-GENE software (Podlich and Cooper, 1998; Wang et al., 2003). A total of 2500 molecular markers were distributed uniformly across 10 chromosomes while 315 QTLs were randomly allocated over the 10 chromosomes to simulate one maize (*Zea mays* L.) population. Each QTL and molecular marker was biallelic and the QTL additive values ranged from 0 to 0.5. As QU-GENE uses recombination fraction rather than map distance to calculate the probability of crossover events, recombination between adjacent pairs of markers was set at 0.0906; for two flanking markers, the QTL was either on the first (recombination between the first marker and QTL was equal to 0.0) or on the second (recombination between the first marker and QTL was equal to 0.0906) marker; excluding the recombination fraction between 15 random QTLs and their flanking markers which was set at 0.5, i.e., complete independence (Haldane, 1919), to simulate linkage equilibrium between 5% of the QTLs and their flanking markers. In addition, all two adjacent QTLs were in complete linkage.

Each of the four traits (T1, T2, T3, and T4) was affected by a different number of QTLs: 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 T1 and T2, T1 and T3, T1 and T4, T2 and T3, T2 and T4, and T3 and T4, respectively.

The genotypic value of each plant was generated based on its haplotypes and the QTL effects for each trait. For each trait, the phenotypic value for each of four replications of each plant was obtained from QU-GENE by setting the per-plot heritability of T1, T2, T3 and T4 at 0.4, 0.6, 0.6, and 0.8, respectively.

In this data set, we restricted traits T1, T2 and T3 (trait T4 was not restricted) in PPG-PSI and PPG-ESIM, and we imposed three predetermined proportional gains in both selection indices, i.e., the vector of predetermined proportional gains was $\mathbf{d}' = [7\ {-3}\ 5]$. Matrices $\mathbf{U}'$ and $\mathbf{D}'$ for this data set for each of the seven selection cycles were

$$\mathbf{U}' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ and } \mathbf{D}' = \begin{bmatrix} 5 & 0 & -7 \\ 0 & 5 & 3 \end{bmatrix}$$

Note that the sign and proportion of the expected genetic gain values for T1, T2 and T3 should change according to the breeder's interest. For this reason, the sign and proportion values of the first eigenvectors for each selection cycle should follow the direction given by the breeder's interest on the expected genetic gain per each trait (Cerón-Rojas et al., 2008a, 2008b). Thus, if the expected genetic gain values for T1 and T3 are positive and negative for T2, if it is necessary, the sign of the values of the first eigenvector will be according to this objective, from where the diagonal matrix $\mathbf{F}$ should have a combination of signs and proportional values to achieve predetermined and positive values for T1 and T3 and predetermined and negative values for T2 in the linear transformation $\beta = \mathbf{F}\mathbf{b}_{\text{PPG-ESIM}}$. When the signs and proportion of the values of the first eigenvectors are according to those required by the breeder's interest, then all the values in diagonal matrix $\mathbf{F}$ will be 1s. The same idea should be applied to traits in simulated data set 2 and the real data set.

Note that the inner product of the PPG-ESIM eigenvector is equal to 1; however, the inner product of the PPG-PSI coefficient vector could not be equal to 1. Thus, the best way of comparing the PPG-ESIM results with those from the PPG-PSI is when the PPG-PSI coefficient vector is normalized, (i.e., when the PPG-PSI coefficient vector is equal to $\mathbf{b}_M^{\star} = \mathbf{b}_M / \mathbf{b}_M'\mathbf{b}_M$), otherwise, the product $\mathbf{b}_M'\mathbf{b}_M$ will spuriously increase the PPG-PSI results. In the present study, the PPG-PSI coefficient vector was normalized only for simulated data set 2.

Simulated Data Set 2

A stochastic simulation was performed to compare the results of applying PPG-PSI and PPG- ESIM selection indices to 10 generations of selection. The initial input for the phenotypic and genetic covariance matrices, and therefore the heritability, economic weights and desired gains for four traits, were taken from Akbar et al. (1984), from where the estimated phenotypic and genetic covariance matrices among number of eggs (RL), age at sexual maturity (SM, days), egg weight (EW, kg), and body weight (BW, kg) were used as initial values for the simulation. The predetermined proportional gains for traits RL, SM, and EW were $\mathbf{d}' = [3\ {-1}\ 2]$ (Lin, 2005) in both selection indices, and the economic weights were all equal to 1.

We used a hypothetical genome with 200 independent segregating sites that affected all traits following a full pleiotropic model according to Zhang et al. (2015). A recurrent selection scheme was performed over 10 cycles, where the breeding design considered 200 full-sib progenies with 200 individuals each. The selection was done based on PPG-PSI and PPG-ESIM with a selection intensity of 20%. From each simulation, we estimated the proportional constant associated with the PPG-PSI coefficient (Eq. [3]); the expected genetic gain for each trait (Eq. [4a] and [7a]), the selection response (Eq. [4b] and [7b]), and the accuracy (Eq. [8]) for both selection indices; finally we estimated the ratio (Eq. [8]) to compare the PPG-PSI efficiency vs. the PPG-ESIM efficiency. The results were summarized based on the average of 200 simulations for each scenario. The genomic recombination routines were implemented in C++ linked to R 3.2.0 through Rcpp (R Development Core Team, 2014; Eddelbuettel, 2013).

Real Data Set

This data set was obtained from Manning (1956) and contains four traits measured in cotton (*Gossypium hirsutum* L.). The traits were: number of cotton balls per plant (V1), number of seeds per ball (V2), lint per seed (V3) and total lint yield per plant (V4), evaluated during seven annual selection cycles from 1949 through 1955. Estimates of the phenotypic and genotypic (in parentheses) variance and covariance of the four traits are given in Table 1.

Similar to simulated data set 1, in all selection cycles we used a value of k = 1.75 for making selections using PPG-PSI and PPG-ESIM. The economic weights for V1, V2, V3 and V4 were 1, 1, 1, and 1, respectively, in both indices. In each selection cycle, we estimated the proportional constant associated with the PPG-PSI coefficient (Eq. [3]), the PPG-PSI and PPG-ESIM expected genetic gain per selection cycle for each trait (Eq. [4a] and [7a]), the PPG-PSI and PPG-ESIM selection response (Eq. [4b] and [7b]), the correlation between the net genetic merit $H = \mathbf{w}'\mathbf{a}$ and PPG-PSI and between $H = \mathbf{w}'\mathbf{a}$ and PPG-ESIM (Eq. [8]), and the ratio $\hat{\pi} = \hat{\lambda} / \hat{\rho}_{H,\mathrm{PSI}}$ (Eq. [8]). In this data set, it was not possible to calculate the true correlation between $H = \mathbf{w}'\mathbf{a}$ and PPG-PSI and between $H = \mathbf{w}'\mathbf{a}$ and PPG-ESIM.

We restricted traits V1, V2 and V3 (trait V4 was not restricted) in both indices and we imposed three predetermined proportional gains in both selection indices, i.e., $\mathbf{d}' = [2\ {-5}\ 10]$. Matrices $\mathbf{U}'$ and $\mathbf{D}'$ for this data set for each of the 7 yr were

$$\mathbf{U}' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ and } \mathbf{D}' = \begin{bmatrix} 10 & 0 & -2 \\ 0 & 10 & 5 \end{bmatrix}$$

The $\mathbf{F}$ matrix (Eq. [5b]) was equal to the identity matrix to facilitate comparing PPG-ESIM results to PPG-PSI results.

RESULTS

Simulated Data Set 1

The first part of Table 2 shows the estimated values of the proportional constant associated with PPG-PSI (Eq. [3]), the estimated PPG-PSI expected genetic gains per selection cycle for each trait (Eq. [4a]), the estimated PPG-PSI selection responses (Eq. [4b]), the correlations between the net genetic merit ($H = \mathbf{w}'\mathbf{a}$) and PPG-PSI (Eq. [8]), and the calculated true correlation between $H = \mathbf{w}'\mathbf{a}$ and PPG-PSI (λ and $\rho_{H,\mathrm{PSI}}$) for three restricted traits T1, T2 and T3 (trait T4 was not restricted) in seven selection cycles.

The second part of Table 2 shows the estimated values of the ratio $\hat{\pi} = \hat{\lambda} / \hat{\rho}_{H,\mathrm{PSI}}$ (Eq. [8]), the estimated PPG-ESIM expected genetic gains per selection cycle for each trait (Eq. [7a]), the estimated PPG-ESIM selection responses (Eq. [7b]), the correlations between the net genetic merit $H = \mathbf{w}'\mathbf{a}$ and PPG-ESIM (Eq. [8]), and the calculated true correlation between $H = \mathbf{w}'\mathbf{a}$ and PPG-ESIM (λ and $\rho_{H,\mathrm{PSI}}$) for three restricted traits T1, T2 and T3 (trait T4 was not restricted) in seven selection cycles.

In all selection cycles, the estimated values of the proportional constant associated with PPG-PSI were positive (Table 2); this means that all PPG-PSI results were valid. The averages of the estimated PPG-PSI expected genetic gains per selection cycle for traits T1, T2 and T3 were 5.20, –2.23 and 3.72, respectively, while the estimated PPG-ESIM expected genetic gains per selection cycle for traits T1, T2 and T3 were 7.74, –2.35 and 2.19, respectively.

Table 1. Real data set. Estimates of the phenotypic and genotypic (in parentheses) variances $\hat{S}^2_{V_1}$, $\hat{S}^2_{V_2}$, $\hat{S}^2_{V_3}$ and $\hat{S}^2_{V_4}$ and covariances $\hat{S}_{V_1V_2}$, $\hat{S}_{V_1V_3}$, $\hat{S}_{V_1V_4}$, $\hat{S}_{V_2V_3}$, $\hat{S}_{V_2V_4}$ and $\hat{S}_{V_3V_4}$ for four cotton traits: number of cotton balls per plant (V1), number of seeds per ball (V2), lint per seed (V3), and total lint yield per plant (V4) in each of seven annual selection cycles (extracted from Manning, 1956).

Year	$\hat{S}^2_{V_1}$	$\hat{S}^2_{V_2}$	$\hat{S}^2_{V_3}$	$\hat{S}^2_{V_4}$	$\hat{S}_{V_1V_2}$	$\hat{S}_{V_1V_3}$	$\hat{S}_{V_1V_4}$	$\hat{S}_{V_2V_3}$	$\hat{S}_{V_2V_4}$	$\hat{S}_{V_3V_4}$
1949	7.298	0.927	0.048	14.618	–0.121	0.09	9.848	–0.087	0.046	0.316
	(4.619)	(0.082)	(0.028)	(10.604)	(–0.058)	(–0.074)	(6.854)	(–0.060)	(0.382)	(0.310)
1950	4.590	1.259	0.075	4.597	0.124	–0.020	2642	–0.077	0.075	0.244
	(0.0)	(0.436)	(0.0561)	(0.871)	(–0.088)	(0.081)	(–0.068)	(–0.051)	(0.040)	(0.161)
1951	1.028	2.157	0.066	2.772	0.566	–0.120	1.098	–0.120	0.915	0.077
	(0.273)	(1.542)	(0.055)	(0.613)	(0.659)	(–0.110)	(0.272)	(–0.101)	(0.712)	(0.025)
1952	6.717	1.253	0.1228	7.591	–0.183	–0.183	5.470	–0.082	0.102	0.261
	(3.606)	(0.452)	(0.105)	(3.671)	(–0.362)	(–0.240)	(5.578)	(–0.073)	(–0.386)	(0.165)
1953	0.785	0.689	0.0374	1.963	–0.050	–0.078	0.071	–0.054	0.077	–0.097
	(0.0)	(0.142)	(0.025)	(0.0)	(–0.192)	(–0.078)	(–1.079)	(–0.049)	(–0.169)	(–0.123)
1954	1.801	0.903	0.0322	2.489	–0.176	–0.070	1.691	–0.007	0.274	0.043
	(0.218)	(0.536)	(0.0119)	(0.397)	(0.334)	(–0.067)	(0.257)	(–0.116)	(0.225)	(–0.001)
1955	0.582	1.315	0.0404	1.535	0.004	–0.015	0.650	–0.077	0.555	–0.009
	(0.118)	(0.338)	(0.0139)	(0.505)	(0.149)	(–0.023)	(0.138)	(–0.068)	(0.515)	(–0.033)

Then, according to the vector of predetermined proportional gains, $\mathbf{d}' = [7\ -3\ 5]$, the estimated PPG-ESIM expected genetic gains per selection cycle for T1 and T2 were closer to the predetermined proportional gains (7.74 and −2.35) than the estimated PPG-PSI expected genetic gains (5.20 and −2.23) per selection cycle for T1 and T2 in all selection cycles; however, the estimated PPG-PSI expected genetic gain per selection cycle for T3 (3.72) was better than that of PPG-ESIM (2.19) (Table 2).

The average values for the estimated PPG-PSI selection response, the correlation between $H = \mathbf{w}'\mathbf{a}$ and PPG-PSI, and the calculated true correlation between $H = \mathbf{w}'\mathbf{a}$ and PPG-PSI were 12.58, 0.77 and 0.74, respectively. The average values for the estimated PPG-ESIM selection response, the correlation between $H = \mathbf{w}'\mathbf{a}$ and PPG-ESIM, and the calculated true correlation between $H = \mathbf{w}'\mathbf{a}$ and PPG- ESIM were 14.79, 0. 996 and 1.0, respectively, while the average value of the ratio $\hat{\pi} = \hat{\lambda} / \hat{\rho}_{H,\mathrm{PSI}}$ (Eq. [8]) was 1.30. This last result indicates that PPG-ESIM efficiency was 30% higher than PPG-PSI efficiency in each selection cycle. In addition, note that the estimated and true correlations between $H = \mathbf{w}'\mathbf{a}$ and PPG-PSI were very similar (Table 2); the same was true for the estimated and true correlations between $H = \mathbf{w}'\mathbf{a}$ and PPG-ESIM (Table 2). This indicates that there was no difference between the estimated and true correlations in both indices, but in all selection cycles PPG-ESIM accuracy was higher than PPG-PSI accuracy.

Simulated Data Set 2

The upper part of Table 3 shows the estimated values of the proportional constants associated with PPG-PSI, the PPG-PSI expected genetic gains per selection cycle for each trait, the PPG-PSI selection responses, the correlations between the net genetic merit and PPG-PSI, and the calculated true correlations between net genetic merit and index PPG-PSI for traits RL, SM, and EW (trait BW was not restricted) in 10 selection cycles.

The lower part of Table 3 shows the estimated values of $\hat{\pi} = \hat{\lambda} / \hat{\rho}_{H,\mathrm{PSI}}$, the PPG-ESIM expected genetic gains per selection cycle for each trait, the PPG-ESIM selection responses, the correlations between the net genetic merit and PPG-ESIM, and the calculated true correlations between the net genetic merit and PPG-ESIM for three restricted traits: RL, SM, and EW (trait BW was not restricted) in 10 selection cycles.

In all selection cycles, the estimated values of the proportional constant associated with PPG-PSI were positive and the average PPG-PSI expected genetic gains per selection cycle for traits RL, SM and EW were 1.68, −0.56 and 1.12, respectively, while the average PPG-ESIM expected genetic gains per selection cycle for traits RL, SM, and EW were 0.71, −2.91 and 6.66, respectively.

Table 2. Simulated data set 1. Estimated values of the Predetermined Proportional Gains Phenotypic Selection Index (PPG-PSI) and the Predetermined Proportional Gains Eigen Selection Index Method (PPG-ESIM) for the proportional constant associated with PPG-PSI (theta value, $\hat{\theta}$), the expected genetic gain per selection cycle for each trait, the selection response ($\hat{R}_{\mathrm{PSI}}$ and $\hat{R}_{\mathrm{ESIM}}$), the estimated ($\hat{\rho}_{H,\mathrm{PSI}}$ and $\hat{\lambda}$) and true ($\rho_{H,\mathrm{PSI}}$ and λ) correlation between the net genetic merit and PPG-PSI and PPG-ESIM and the ratio $\hat{\pi} = \hat{\lambda} / \hat{\rho}_{H,\mathrm{PSI}}$, for restricted traits T1, T2 and T3, using d′= [7 −3 5] in seven selection cycles.

	PPG-PSI estimated values							
	Theta	Expected genetic gain				Response	Correlation	True correlation
Cycle	$\hat{\theta}$	T1	T2	T3	T4	$\hat{R}_{\mathrm{PSI}}$	$\hat{\rho}_{H,\mathrm{PSI}}$	$\rho_{H,\mathrm{PSI}}$
1	4.53	6.41	−2.75	4.58	1.50	15.23	0.78	0.75
2	3.93	5.89	−2.52	4.21	1.77	14.39	0.81	0.78
3	3.35	5.48	−2.35	3.91	1.45	13.18	0.80	0.77
4	2.55	4.76	−2.04	3.4	1.36	11.56	0.70	0.70
5	2.87	5.08	−2.18	3.63	1.28	12.16	0.76	0.74
6	2.19	4.39	−1.88	3.14	1.36	10.77	0.74	0.74
7	2.20	4.41	−1.89	3.15	1.30	10.75	0.77	0.72
Average	3.09	5.20	−2.23	3.72	1.43	12.58	0.77	0.74
	PPG-ESIM estimated values							
	Ratio	Expected genetic gain				Response	Correlation	True correlation
Cycle	$\hat{\pi}$	T1	T2	T3	T4	$\hat{R}_{\mathrm{ESIM}}$	$\hat{\lambda}$	λ
1	1.27	8.51	−3.79	2.98	1.01	16.44	0.99	1.0
2	1.21	8.31	−2.47	2.67	0.72	15.31	0.98	1.0
3	1.25	7.56	−2.37	2.14	1.17	12.56	1.00	1.0
4	1.43	8.48	−2.40	1.88	0.26	17.07	1.00	1.0
5	1.30	7.90	−1.99	2.08	0.60	15.15	0.99	1.0
6	1.34	7.25	−1.72	1.57	0.40	13.83	0.99	1.0
7	1.32	6.19	−1.69	2.03	0.31	13.17	1.02	1.0
Average	1.30	7.74	−2.35	2.19	0.64	14.79	0.99	1.0

Table 3. Simulated data set 2. Estimated values of the Predetermined Proportional Gains Phenotypic Selection Index (PPG-PSI) and the Predetermined Proportional Gains Eigen Selection Index Method (PPG-ESIM) for the proportional constant associated with PPG-PSI (theta value, $\hat{\theta}$), the expected genetic gain per selection cycle for each trait, the selection response ($\hat{R}_{PSI}$ and $\hat{R}_{ESIM}$), the estimated ($\hat{\rho}_{H,PSI}$ and $\hat{\lambda}$) and true ($\rho_{H,PSI}$ and λ) correlation between the net genetic merit and PPG-PSI and PPG-ESIM and the ratio $\hat{\pi} = \hat{\lambda} / \hat{\rho}_{H,PSI}$, for three restrictive traits RL, SM, and EW, using d′= [3 –1 2] in 10 selection cycles.

	PPG-PSI estimated values							
	Theta	Expected genetic gain				Response	Correlation	True correlation
Cycle	$\hat{\theta}$	RL†	SM	EW	BW	$\hat{R}_{PSI}$	$\hat{\rho}_{H,PSI}$	$\rho_{H,PSI}$
1	18.80	1.64	–0.55	1.09	58.30	60.49	0.42	0.41
2	19.48	1.74	–0.58	1.16	56.98	59.30	0.42	0.42
3	18.75	1.69	–0.56	1.12	56.32	58.56	0.41	0.40
4	18.33	1.63	–0.54	1.09	55.36	57.54	0.41	0.41
5	19.34	1.75	–0.58	1.17	55.87	58.21	0.42	0.41
6	17.38	1.63	–0.54	1.09	54.28	56.46	0.41	0.42
7	18.03	1.71	–0.57	1.14	53.99	56.26	0.42	0.42
8	17.58	1.70	–0.57	1.13	52.93	55.19	0.41	0.41
9	17.55	1.72	–0.57	1.15	52.41	54.70	0.41	0.41
10	16.28	1.64	–0.55	1.09	51.04	53.22	0.41	0.40
Average	18.15	1.68	–0.56	1.12	54.75	56.99	0.41	0.41
	PPG-ESIM estimated values							
	Ratio	Expected genetic gain				Response	Correlation	True correlation
Cycle	$\hat{\pi}$	RL	SM	EW	BW	$\hat{R}_{ESIM}$	$\hat{\lambda}$	λ
1	1.70	0.85	–3.15	7.45	78.23	196.80	0.71	0.57
2	1.68	0.73	–3.09	6.99	74.90	186.37	0.70	0.57
3	1.70	0.75	–3.03	6.79	74.27	182.57	0.71	0.57
4	1.70	0.52	–2.85	6.83	73.38	180.22	0.70	0.57
5	1.67	0.47	–2.92	6.79	72.57	179.45	0.70	0.57
6	1.69	0.72	–2.96	6.58	71.87	176.51	0.70	0.57
7	1.70	0.84	–2.97	6.55	71.03	175.07	0.71	0.57
8	1.69	0.75	–2.87	6.32	69.25	169.77	0.70	0.57
9	1.68	0.63	–2.68	6.31	68.76	166.46	0.70	0.57
10	1.67	0.85	–2.56	5.99	66.12	157.02	0.68	0.56
Average	1.69	0.71	–2.91	6.66	72.04	177.02	0.70	0.57

† RL, number of eggs; SM; age at sexual maturity; EW, egg weight; BW, body weight.

Real Data Set

The first part of Table 4 shows the estimated values of the proportional constant associated with PPG-PSI (Eq. [3]), the estimated PPG-PSI expected genetic gain per selection cycle for each trait (Eq. [4a]), the estimated PPG-PSI selection response (Eq. [4b]), and the correlation between the net genetic merit ($H = \mathbf{w}'\mathbf{a}$) and PPG-PSI (Eq. [8]) for the restricted traits: number of cotton balls per plant (V1), number of seeds per ball (V2) and lint per seed (V3) (trait total lint yield per plant, V4, was not restricted) using $\mathbf{d}' = [2\ –5\ 10]$ in seven annual selection cycles (from 1949 through 1955).

The second part of Table 4 shows the estimated values of the ratio $\hat{\pi} = \hat{\lambda} / \hat{\rho}_{H,PSI}$ (Eq. [8]), the estimated PPG-ESIM expected genetic gain per selection cycle for each trait (Eq. [7a]), the estimated PPG-ESIM selection responses (Eq. [7b]), and the estimated correlations between the net genetic merit $H = \mathbf{w}'\mathbf{a}$ and PPG-ESIM (Eq. [8]) for the restricted traits: number of cotton balls per plant (V1), number of seeds per ball (V2) and lint per seed (V3) (trait total lint yield per plant, V4, was not restricted) using $\mathbf{d}' = [2\ –5\ 10]$ in seven annual selection cycles (from 1949 through 1955).

Note that for this data set, the estimated values of the proportional constant associated with PPG-PSI (Eq. [3]) were negative for 1951, 1952, and 1953 (Table 4). This means that the PPG-PSI results were valid only for 1949, 1950, 1954, and 1955. Valid averages for 4 yr (1949, 1950, 1954, and 1955) of the estimated PPG-PSI expected genetic gain per selection cycle for each trait (V1, V2 and V3) were 0.039, −0.096 and 0.192, respectively.

Similar to the PPG-PSI index, note that for 1949, 1952, and 1953, the estimated eigenvalues were higher than 1, $\hat{\lambda} > 1$ (Table 4). This is because although the estimated matrix $\hat{\mathbf{P}}$ was positive definite (all eigenvalues positive), the estimated matrix $\hat{\mathbf{C}}$ was not positive semidefinite (no negative eigenvalues), i.e., some $\hat{\mathbf{C}}$ eigenvalues for 1949, 1952, and 1953 were negative. Thus the PPG-ESIM results were valid only for 1950, 1951, 1954, and 1955. The PPG-ESIM valid averages for the 4 yr (1950, 1951, 1954, and 1955) of the estimated PPG-ESIM expected

Table 4. Real data set. Estimated values of the Predetermined Proportional Gains Phenotypic Selection Index (PPG-PSI) and the Predetermined Proportional Gains Eigen Selection Index Method (PPG-ESIM) for the proportional constant associated with PPG-PSI (theta value, $\hat{\theta}$), the expected genetic gain per selection cycle for each trait, the selection response ($\hat{R}_{PSI}$ and $\hat{R}_{ESIM}$), the estimated correlation ($\hat{\rho}_{H,PSI}$ and $\hat{\lambda}$) between the net genetic merit and PPG-PSI and PPG-ESIM, and the ratio ($\hat{\pi} = \hat{\lambda} / \hat{\rho}_{H,PSI}$) for four real cotton traits: number of cotton balls per plant (V1), number of seeds per ball (V2), lint per seed (V3), and total lint yield per plant (V4) in each of seven annual selection cycles (extracted from Manning, 1956). Traits V1, V2 and V3 were restricted with d′= [2 −5 10] in all selection cycles.

	PPG-PSI estimated values						
	Theta	Expected genetic gain				Response	Correlation
Year	$\hat{\theta}$	V1	V2	V3	V4	$\hat{R}_{PSI}$	$\hat{\rho}_{H,PSI}$
1949	0.0302	0.053	−0.132	0.263	3.346	3.531	0.37
1950	0.0058	0.034	−0.085	0.169	0.947	1.065	0.49
1951	−0.0001	−0.002	0.004	−0.008	0.429	0.424	0.10
1952	−0.0011	−0.002	0.004	−0.008	4.163	4.158	0.57
1953	−0.0003	−0.001	0.002	−0.004	2.449	2.446	0.78
1954	0.0063	0.064	−0.160	0.320	0.381	0.605	0.22
1955	0.0002	0.003	−0.007	0.014	0.460	0.470	0.18
Valid Average†	0.011	0.039	−0.096	0.192	1.284	1.481	0.315
	PPG-ESIM estimated values						
	Ratio	Expected genetic gain				Response	Lambda
Year	$\hat{\pi}$	V1	V2	V3	V4	$\hat{R}_{ESIM}$	$\hat{\lambda}$
1949	7.14	3.139	0.168	0.142	4.861	19.092	2.64
1950	1.35	−0.038	−0.006	0.141	0.710	2.426	0.66
1951	5.50	0.270	0.718	0.053	0.635	2.148	0.55
1952	3.88	3.536	−0.263	0.109	2.384	10.722	2.21
1953	2.88	1.324	0.184	0.154	−0.028	5.493	2.25
1954	3.18	0.079	−0.066	0.146	0.332	1.444	0.70
1955	5.56	0.187	0.707	−0.041	0.684	2.685	1.00
Valid Average‡	2.910	0.125	0.338	0.075	0.590	2.176	0.728

† Averages were obtained only for 1949, 1950, 1954, and 1955. Values where $\hat{\theta}$ was negative were not considered.

‡ Averages were obtained only for 1950, 1951, 1954, and 1955. Values where $\hat{\lambda} > 1$ were not considered.

genetic gain per selection cycle for each trait (V1, V2 and V3) were 0.125, 0.338 and 0.075, respectively.

According to the vector of predetermined proportional gains $\mathbf{d}' = [2 -5 10]$, the estimated expected genetic gains per selection cycle for the three traits in both selection indices were not precise. These results may be due to the lowest phenotypic and genotypic variance and covariance (Table 1).

The valid average values for the estimated PPG-PSI selection responses and the correlations between $H = \mathbf{w}'\mathbf{a}$ and PPG-PSI were 1.481 and 0.315, respectively. On the other hand, the valid average value for the estimated PPG-ESIM selection responses was 2.176 and its correlations with $H = \mathbf{w}'\mathbf{a}$ was 0.728 (Table 4). The valid average value of the ratios $\hat{\pi} = \hat{\lambda} / \hat{\rho}_{H,PSI}$ (Eq. [8]) was 2.91. This last result indicates that PPG-ESIM efficiency was almost twice as high as PPG-PSI efficiency in each selection cycle.

DISCUSSION

Results for two simulated data sets and one real data set showed that PPG-ESIM is more efficient than PPG-PSI because in the three data sets, the estimated PPG-ESIM selection responses and the estimated correlations between the net genetic merit and PPG-ESIM were higher than the estimated PPG-PSI selection responses and the estimated correlations between the net genetic merit and PPG-PSI.

The results obtained using the simulated data set for comparing PPG-PSI efficiency vs. PPG-ESIM efficiency showed that the average PPG-ESIM expected genetic gains for two of three restricted traits were closer to the predetermined proportion than the PPG-PSI expected genetic gains. The PPG-PSI and PPG-ESIM expected genetic gains obtained using the real data set were less precise than those obtained using the two simulated data sets: however, PPG-ESIM accuracies were also higher than PPG-PSI accuracies for all restricted traits in all seven selection cycles.

As the objective of any PSI is to predict the net genetic merit ($H = \mathbf{w}'\mathbf{a}$), the correlation between PPG-PSI and H, and between PPG-ESIM and H, should be the maximum value possible (Hazel, 1943). Thus, as the correlation between PPG-ESIM and H was always higher than the correlation between PPG-PSI and H, PPG-ESIM is a better predictor of H than PPG-PSI from this point of view. In addition, the ratio $\rho_{H,ESIM} / \rho_{H,PSI}$ (Eq. [8]) was higher than 1 for the three data sets.

Why PPG-ESIM Accuracies were Higher than PPG-PSI Accuracies for the Three Data Sets

Note that $\rho_{H,\text{PSI}}$ was maximized only with respect to the vector of Eq. [2a] ($\mathbf{b}_M$), while $\rho_{H,\text{PSI}}$ was maximized with respect to the vector of Eq. [5a] ($\mathbf{b}_{\text{PPG-ESIM}}$) and with respect to the estimated vector of economic values $\mathbf{w}$, which, in the PPG-ESIM context, can be written as $\mathbf{w}_E = \mathbf{C}^{-1}[\lambda\mathbf{Pb}_{\text{PPG-ESIM}} + \mathbf{Mv}]$ (Eq. [6]). We believe this is the main reason why in all cases $\hat{\rho}_{H,\text{ESIM}} > \hat{\rho}_{H,\text{PSI}}$.

PPG-ESIM Estimated Economic Weights

The economic weight for each trait depends on the amount by which profit may be expected to increase for each unit of improvement in that trait. In the context of animal breeding, these values may vary from breed to breed or from region to region within the same breed and they may change, even while a breeding program is in progress, if permanent shifts in market demand occur (Hazel, 1943). In the most favorable case, when we have complete information, multiple regression techniques can be used to determine the economic weights. In this case, the economic weights are unbiased, but usually have fairly large sampling errors. For some traits, however, economic information is lacking or only partially available. In these circumstances, economic weights are intelligent guesses rather than accurate estimates (Vandepitte and Hazel, 1977).

Vandepitte and Hazel (1977) and Smith (1983) ran computer simulations to evaluate the economic weight sampling error for five traits and concluded that the effects on the SPSI for errors of up to 50% (in one trait at a time) had small effects on selection efficiency, but that large changes (±200%) in some traits caused substantial losses in SPSI efficiency. Based on economic theory, Melton et al. (1979) proposed a profit function to estimate the economic weight; however, Goddard (1983) found inconsistencies in the method of Melton et al. (1979). Magnussen (1990) used an approach for estimating economic weight that is similar to the one used in Eq. [6].

In the PPG-ESIM context, we can estimate $\mathbf{w}$ as one linear combination of the first eigenvector and the first eigenvalue. Note that in the canonical correlation theory (Anderson, 2003), there are two linear combinations (similar to the net genetic merit $H = \mathbf{w}'\mathbf{a}$ and to the linear phenotypic selection index $I = \mathbf{b}'\mathbf{p}$), and in each one, the vectors of coefficients are unknown and thus have to be estimated; the same argument is valid for the PPG-ESIM theory. Finally, note that the estimated economic weights are valid only for the PPG-ESIM, that is, it is not a general method to estimate economic weights for any linear phenotypic selection index.

Importance of the PPG-PSI

The PPG-PSI allows imposing restrictions on the expected genetic gain values to make some traits change their mean values based on a predetermined level, while the rest of the traits remain without restrictions. By this reasoning, the PPG-PSI should be the basic tool for selecting individuals as parents of the next generation. We think that the Lande and Thompson (1990) index, the molecular eigen selection index (Cerón-Rojas et al., 2008b), and the Dekkers (2007) index, which use molecular marker information for making selections (the first two combine quantitative trait loci information with phenotypic information, whereas the third combines genomic estimated breeding values with phenotypic information), should be adapted to the PPG-PSI or the PPG-ESIM. The same should be done with the genomic selection indices of Togashi et al. (2011) and Ceron-Rojas et al. (2015).

Importance of the PPG-ESIM

The PPG-ESIM does not require a proportional constant and, due to the properties associated with eigen analysis, it is possible to use the theory of similar matrices to change the direction of the eigenvector values without affecting the correlation between the PPG-ESIM and the net genetic merit (Harville, 1997), which helps to eliminate the problem of shifting the population means in the opposite direction to the predetermined desired direction. The accuracy of PPG-ESIM is always higher than PPG-PSI accuracy because the PPG-ESIM accuracy is not affected by the increasing number of restricted traits.

Hayes and Hill (1980) have shown that when the eigenvalues of matrix $\hat{\mathbf{P}}^{-1}\hat{\mathbf{C}}$ are negative or greater than one, matrix $\hat{\mathbf{P}}$ is not positive definite (all eigenvalues positive) or $\hat{\mathbf{C}}$ is not positive semidefinite (no negative eigenvalues); thus the estimated $\hat{\mathbf{P}}$ and/or $\hat{\mathbf{C}}$ matrix values are wrong. This result is valid in the context of matrix $\hat{\mathbf{T}} = \hat{\mathbf{K}}\hat{\mathbf{P}}^{-1}\hat{\mathbf{C}}$, which is used to obtain the PPG-PSI and PPG-ESIM vectors of coefficients (Eq. [2a] and [5a]). We believe this is another important characteristic of the PPG-ESIM because the eigenvalues of matrix $\hat{\mathbf{T}}$ can be used to check the values of matrices $\hat{\mathbf{P}}$ and $\hat{\mathbf{C}}$.

In addition, note that matrices $\mathbf{K} = [\mathbf{I} - \mathbf{Q}]$ and $\mathbf{Q}$ are projectors. That is, matrices $\mathbf{K}$ and $\mathbf{Q}$ are idempotent ($\mathbf{K} = \mathbf{K}^2$ and $\mathbf{Q} = \mathbf{Q}^2$) and unique (Searle, 1966); they are also orthogonal, i.e., $\mathbf{KQ} = \mathbf{QK} = 0$. In Eq. [2a], matrix $\mathbf{Q}$ projects vector $\mathbf{b}$ into a space generated by the columns of matrix $\mathbf{M}$ due to the restriction $\mathbf{M}'\mathbf{b} = 0$ that is introduced when Φ_M is maximized with respect to $\mathbf{b}$, while matrix $\mathbf{K}$ projects vector $\mathbf{b}$ into a space perpendicular to that generated by the columns of matrix $\mathbf{M}$ (Rao, 2002; Cerón-Rojas et al., 2016). That is, the main function of matrix $\mathbf{K}$ is to transform vector $\mathbf{b}$ into vector $\mathbf{b}_M$.

Matrix $\mathbf{K}$ has a similar function in Eq. [5a] but is adapted to the canonical correlation theory (Anderson, 1999, 2003). Thus, note that vectors $\mathbf{p}$ and $\mathbf{a}$ (Eq. [1a]) can be ordered in a new vector $\mathbf{x}$ as $\mathbf{x}' = [\mathbf{p}' \ \mathbf{a}]$, from where the covariance matrix of $\mathbf{x}$ is $\begin{bmatrix} \mathbf{P} & \mathbf{C} \\ \mathbf{C} & \mathbf{C} \end{bmatrix}$. One measure of the association between a linear combination of $\mathbf{p}$ and a linear combination of $\mathbf{a}$ is the canonical correlation (λ_j) value obtained from equation $(\mathbf{P}^{-1}\mathbf{C} - \lambda_j^2\mathbf{I})\beta_j = 0$, where β_j is the jth eigenvector of matrix $\mathbf{P}^{-1}\mathbf{C}$ and λ_j is the jth (j = 1, 2, …, t) canonical correlation value of a linear combination of $\mathbf{p}$ and a linear combination of $\mathbf{a}$. When we maximized Equation Φ with respect to $\mathbf{b}$, $\mathbf{w}$, 0.5λ, 0.5μ, and $\mathbf{v}'$, we introduced matrix $\mathbf{K}$ in equation $(\mathbf{P}^{-1}\mathbf{C} - \lambda_j^2\mathbf{I})\beta_j = 0$, as can be seen in Eq. [5a]. These results simplified the PPG-ESIM vector of coefficients because it does not need a proportional constant and thus is simpler than the PPG-PSI vector of coefficients. Thus, PPG-ESIM efficiency is greater than PPG-PSI efficiency in part because PPG-ESIM vector of coefficients is simpler than PPG-PSI vector of coefficients.

CONCLUSIONS

We proposed a predetermined proportional gains phenotypic selection index based on the eigen selection index method (PPG-ESIM), which does not require a proportional constant. Due to the properties associated with eigen-analysis, it is possible to use the theory of similar matrices to change the direction of the eigenvector values without affecting the correlation between PPG-ESIM and the net genetic merit. PPG-ESIM uses the first eigenvector for determining the proportion of each trait contributing to PPG-ESIM, and the first eigenvalue in the PPG-ESIM selection response. Results obtained using two simulated data sets and one real data set indicated that, in all cases, PPG-ESIM efficiency was higher than the predetermined proportional gains phenotypic selection index efficiency. We concluded that PPG-ESIM is a good selection index that can be used in any phenotypic selection program.

Acknowledgments

The authors are grateful to México's Consejo Nacional de Ciencia y Tecnología (CONACYT) for partially funding this research.

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

Anderson, T.W. 1999. Asymptotic theory for canonical correlation analysis. J. Multivariate Anal. 70:1–29. doi:10.1006/jmva.1999.1810

Anderson, T.W. 2003. An introduction to multivariate statistical analysis. 3rd ed., John Wiley & Sons, New Jersey.

Baker, R.J. 1974. Selection indexes without economic weights for animal breeding. Can. J. Anim. Sci. 54:1–8. doi:10.4141/cjas74-001

Bulmer, M.G. 1980. The mathematical theory of quantitative genetics. Lectures in Biomathematics. Univ. of Oxford, Clarendon Press, England.

Cerón-Rojas, J.J., J. Crossa, and J. Sahagún-Castellanos. 2016. Statistical sampling properties of the coefficients of three phenotypic selection indices. Crop Sci. 56:51–58. doi:10.2135/cropsci2015.03.0189

Cerón-Rojas, J.J., J. Crossa, J. Sahagún-Castellanos, F. Castillo-González, and A. Santacruz-Varela. 2006. A selection index method based on eigen analysis. Crop Sci. 46:1711–1721. doi:10.2135/cropsci2005.11-0420

Ceron-Rojas, J.J., J. Crossa, V.N. Arief, K. Basford, J. Rutkoski, D. Jarquín, G. Alvarado, Y. Beyene, K. Semagn, and I. DeLacy. 2015. A genomic selection index applied to simulated and real data. Genes Genomes Genetics 5:2155–2164.

Cerón-Rojas, J.J., J. Sahagún-Castellanos, F. Castillo-González, A. Santacruz-Varela, and J. Crossa. 2008a. A restricted selection index method based on eigenanalysis. J. Agric. Biol. Environ. Stat. 13(4):440–457. doi:10.1198/108571108X378911

Cerón-Rojas, J.J., J. Sahagún-Castellanos, F. Castillo-González, A. Santacruz-Varela, I. Benítez-Riquelme, and J. Crossa. 2008b. A molecular selection index method based on eigenanalysis. Genetics 180:547–557. doi:10.1534/genetics.108.087387

Dekkers, J.C.M. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. J. Anim. Breed. Genet. 124:331–341. doi:10.1111/j.1439-0388.2007.00701.x

Eddelbuettel, D. 2013. Seamless R and C++ integration with Rcpp. Springer, New York.

Goddard, M.E. 1983. Selection indices for non-linear profit functions. Theor. Appl. Genet. 64:339–344. doi:10.1007/BF00274177

Haldane, J.B.S. 1919. The combination of linkage values and the calculation of distance between the loci of linked factors. J. Genet. 8:299–309.

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

Harville, D.A. 1997. Matrix algebra from a statistician's perspective. Springer, New York. doi:10.1007/b98818

Hayes, J.F., and W.G. Hill. 1980. A reparameterization of a genetic selection index to locate its sampling properties. Biometrics 36(2):237–248. doi:10.2307/2529975

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

Itoh, Y., and Y. Yamada. 1987. Comparisons of selection indices achieving predetermined proportional gains. Genet. Sel. Evol. 19(1):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.

Lin, C.Y. 2005. A simultaneous procedure for deriving selection indexes with multiple restrictions. J. Anim. Sci. 83:531–536.

Magnussen, S. 1990. Selection index: Economic weights for maximum simultaneous genetic gains. Theor. Appl. Genet. 79:289–293. doi:10.1007/BF01186069

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

Manning, H.L. 1956. Yield improvement from a selection index technique with cotton. Heredity 10:303–322. doi:10.1038/hdy.1956.29

Melton, B.E., E.O. Heady, and R.L. Willham. 1979. Estimation of economic values for selection indices. Anim. Prod. 28:279–286. doi:10.1017/S0003356100023199

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

R Development Core Team. 2014. R: A language and environment for statistical computing. R Development Core Team, Vienna, Austria.

Rao, C.R. 2002. Linear statistical inference and its applications. 2nd ed. John Wiley & Sons, New York.

Searle, S.R. 1966. Matrix algebra for the biological sciences. John Wiley & Sons, New York.

Smith, C. 1983. Effects of changes in economic weights on the efficiency of index selection. J. Anim. Sci. 56(5):1057–1064.

Smith, H.F. 1936. A discriminant function for plant selection In: Papers on quantitative genetics and related topics. North Carolina State College Dep. of Genetics, Raleigh, NC. p. 466–476.

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

Togashi, K., C.Y. Lin, and T. Yamazaki. 2011. The efficiency of genome-wide selection for genetic improvement of net merit. J. Anim. Sci. 89:2972–2980. doi:10.2527/jas.2009-2606

Vandepitte, W.M., and L.N. Hazel. 1977. The effect of errors in the economic weights on the accuracy of selection indexes. Ann. Genet. Sel. Anim. 9:87–103.

Wang, J., M. van Ginkel, D.W. Podlich, G. Ye, R. Trethowan, W. Pfeiffer, I.H. DeLacy, M. Cooper, and S. Rajaram. 2003. Comparison of two breeding strategies by computer simulation. Crop Sci. 43:1764–1773. doi:10.2135/cropsci2003.1764

Zhang, Z., X. Li, X. Ding, J. Li, and Q. Zhang. 2015. GPOP-SIM: A simulation tool for whole-genome genetic data. BMC Genet. doi:10.1186/s12863-015-0173-4