

A genomic selection index applied to simulated and real data

J. Jesus Ceron-Rojas^{*}, José Crossa^{*1}, Vivi N. Arief[§], Kaye Basford[§], Jessica Rutkoski[†], Diego

Jarquín[‡], Gregorio Alvarado^{*}, Yoseph Beyene^{**}, Kassa Semagn^{**}, and Ian DeLacy[§]

^{*}Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT).

Apdo. Postal 6-641, 06600, México DF, México;

[§]The University of Queensland, School of Agriculture and Food Sciences, St Lucia, QLD 4072,

Brisbane, Australia;

[†]International programs of the College of Agriculture and Life Sciences, Cornell Univ., 252

Emerson Hall, Ithaca NY 14853 and CIMMYT;

[‡]Department of Agronomy and Horticulture, University of Nebraska, 363 Keim Hall, Lincoln,

NE 68583, USA;

^{**}Global Maize Program, CIMMYT, P. O. Box 1041, Village Market 00621, Nairobi, Kenya.

¹ Corresponding author: José Crossa, email: j.crossa@cgiar.org

File S1

Supplemental Materials

Criteria for comparing PSI efficiency vs GSI efficiency

Besides Equation (8), we used two additional criteria to compare the efficiency of PSI vs GSI.

The first criterion was based on the average of the top 10% of the estimators of the PSI,

$\mathbf{PSI}_E = \sum_{q=1}^t \hat{b}_q \mathbf{p}_q$, and $\mathbf{GSI}_E = \sum_{q=1}^t w_q \hat{\gamma}_{lq}$ values (both vectors of size $g \times 1$), where \hat{b}_q is the q^{th}

element of vector $\hat{\mathbf{b}}' = \mathbf{w}' \hat{\mathbf{C}} \mathbf{P}^{-1} = [\hat{b}_1 \quad \hat{b}_2 \quad \dots \quad \hat{b}_t]$ ($\hat{\mathbf{b}}$ is an estimator of $\mathbf{b} = \mathbf{P}^{-1} \mathbf{C} \mathbf{w}$) and \mathbf{p}_q is a

$g \times 1$ vector of the q^{th} trait's phenotypic values; \mathbf{GSI}_E was defined in Equation (6). From the

\mathbf{PSI}_E and \mathbf{GSI}_E values, we obtained the percentage of change in the predicted values of the net

genetic merit ($\mathbf{H} = \mathbf{w}' \mathbf{a}$) from one cycle to the next using the formulae:

$$\lambda_p = 100 \left(\frac{\mathbf{PSI}_{E_{i+1}}}{\mathbf{PSI}_{E_i}} - 1 \right) \text{ and } \lambda_g = 100 \left(\frac{\mathbf{GSI}_{E_{i+1}}}{\mathbf{GSI}_{E_i}} - 1 \right) \quad (\text{S1})$$

where $\mathbf{PSI}_{E_{i+1}}$ and $\mathbf{GSI}_{E_{i+1}}$ denote the average of the top 10% \mathbf{PSI}_E and \mathbf{GSI}_E values in the

$(i+1)^{\text{th}}$ cycle ($i=1,2,\dots,s,s=\text{number of selection cycles}$), and \mathbf{PSI}_{E_i} and \mathbf{GSI}_{E_i} denote the

average of the top 10% \mathbf{PSI}_E and \mathbf{GSI}_E values in the i^{th} cycle. Then when λ_p and λ_g are

positive, the change in the average of the top 10% \mathbf{PSI}_E and \mathbf{GSI}_E values from cycle i to cycle

$i+1$ will be positive; when λ_p and λ_g are negative, the change in the average of the top 10%

\mathbf{PSI}_E and \mathbf{GSI}_E values from cycle i to cycle $i+1$ will be negative, and when λ_p and λ_g are

zero, no changes will be observed in the average of the top 10% \mathbf{PSI}_E and \mathbf{GSI}_E values from

cycle i to cycle $i+1$.

The second criterion was based on the Technow *et al.* (2013) inequality which, in the context of PSI and GSI, can be written as

$$L_{GSI} < \frac{\rho_{H,GSI}}{h_{PSI}} < L_{PSI} \quad (S2)$$

where L_{GSI} and L_{PSI} denote the time required for GSI and PSI, respectively, to complete one selection cycle; in the case of maize, GSI requires 1.5 years and PSI requires 4.0 years (Beyene *et al.* 2015); $\rho_{H,GSI}$ is the correlation or accuracy between GSI and the net genetic merit ($H = \mathbf{w}'\mathbf{a}$)

and h_{PSI} is the square root of the heritability of PSI, which can be denoted as $h_{PSI} = \sqrt{\frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{b}'\mathbf{P}\mathbf{b}}}$ (Lin

and Allaire 1977). Then, according to Technow *et al.* (2013), when, in practice, Equation (S2) is true, GSI will be more efficient than PSI in terms of the units of time, assuming that selection intensity and variance of index is the same for both PSI and GSI.

Proportion of change in the predicted net genetic merit values

Simulated data

Table S1 contains the average of the top 10% values and the proportion of change in the predicted net genetic merit values for \mathbf{PSI}_E and \mathbf{GSI}_E from cycle 1 to cycle 7 selected with a selection intensity of $k = 1.75$. These are the averages of the predicted values of H from the top 10% of individuals, and the average for \mathbf{PSI}_E is higher than the average for \mathbf{GSI}_E in each selection cycle. However, the proportional changes from one cycle to the next of these average values are higher for \mathbf{GSI}_E than for \mathbf{PSI}_E (Table S1). The mean proportion from one cycle to

the next for PSI and GSI can be presented as a percentage, according to Equation (S1). For example, the proportional changes (1.02 and 1.21 from cycle 2 to cycle 3) can be expressed as $\lambda_p = 2\%$ for PSI and as 21% for GSI. These results indicated that, on average, GSI was more efficient than PSI from cycle 2 to cycle 3. In terms of the overall averages of the proportion for \mathbf{PSI}_E and \mathbf{GSI}_E , the average efficiency of GSI was 16% and the average efficiency of PSI was 6%.

Real Data

For data sets 3, 4, 5 and 6, selection was conducted for only one PSI cycle (C0) and three GSI cycles (C0, C1 and C2) of the base population and three sets of markers (Table 1). We estimated the selection response per cycles 0, 1 and 2 (Table 3).

In percentages, the average change in the top 10% values for \mathbf{GSI}_E from cycle 1 to 2 for real data set 3 is $\lambda_g = -1$, a decrease of 1% (Table S2). However, the overall averages for real data sets 3, 5 and 6 were 8%, 23% and 13%, respectively, and only real data set 4 had a negative increase of 5%.

Results presented in Tables S1 and S2 indicate that GSI is a better predictor of $H = \mathbf{w}'\mathbf{a}$ than PSI; for the simulated data set used in this study, GSI was slightly more efficient than PSI (in terms of the average of the top 10% values for \mathbf{PSI}_E and \mathbf{GSI}_E). In the real data sets, we do not have information on \mathbf{PSI}_E except for the training population (cycle 0) but the average values of \mathbf{GSI}_E indicated that the average changes from cycle 0 to 1 and from cycle 1 to 2 (Table S2) are similar to those of the simulated data set (Table S1). These results indicated that the simulated data sets used in this study mimic the real situation very well.

Efficiency of GSI and PSI based on cycle length

Simulated data

Table S3 shows the values of the left-hand side (L_{GSI}) and the estimated values of the right-hand side ($\frac{\rho_{H,GSI}}{h_{PSI}} L_{PSI}$) of the Technow *et al.* (2013) inequality for seven PSI and GSI selection cycles.

According to these results, GSI is a better method than PSI in terms of unit time needed to select for the net genetic merit, because in all cycles, the right-hand side of the Technow *et al.* (2013) inequality formula was higher than the left-hand side.

Real data

According to the Technow *et al.* (2013) inequality formula, GSI was more efficient than PSI for data sets 3, 4, 5 and 6, because the right-hand side of the Technow *et al.* (2013) inequality formula was higher than the left-hand side in all data sets (Table S4).

Table S1. Means and proportional changes from one cycle to the next (and their average over the cycles) of the top 10% of phenotypic selection index (\mathbf{PSI}_E) and genomic selection index (\mathbf{GSI}_E) estimated simulated values. The total averages of \mathbf{PSI}_E and \mathbf{GSI}_E values are shown in the last line of the table. To convert the proportional changes to percentages, subtract 1 from the proportional change value and multiply the result by 100.

Cycle	Means		Proportional changes ^a	
	\mathbf{PSI}_E	\mathbf{GSI}_E	\mathbf{PSI}_E	\mathbf{GSI}_E
1	120.5	38.4	--	--
2	143.3	48.6	1.19	1.27
3	146.2	58.7	1.02	1.21
4	161.3	67.8	1.10	1.16
5	170.6	76.3	1.06	1.13
6	169.8	84.1	1.00	1.10
7	169.1	92.3	1.00	1.10
Average	154.4	66.6	1.06	1.16

^a Obtained as $\frac{\mathbf{PSI}_{E_{i+1}}}{\mathbf{PSI}_{E_i}}$ and $\frac{\mathbf{GSI}_{E_{i+1}}}{\mathbf{GSI}_{E_i}}$ $i = 1, 2, \dots, 7$.

Table S2. Means and proportional changes from one cycle to the next (and their average over the three cycles) of the top 10% estimated genomic selection index (\mathbf{GSI}_E) real values for different sets of markers and unobserved QTLs in linkage disequilibrium for four traits: grain yield (GY), ear height (EHT), plant height (PHT) and anthesis day (AD) in four real data sets: 3, 4, 5 and 6. To convert the proportional changes to percentages, subtract 1 from the proportional change value and multiply the result by 100.

Cycle	\mathbf{GSI}_E means				\mathbf{GSI}_E proportional changes ^b			
	3	4	5	6	3	4	5	6
0	7.0	6.3	6.9	4.0	--	--	--	--
1	8.1	6.8	7.1	4.1	1.16	1.08	1.03	1.03
2	8.0	5.6	10.1	5.0	0.99	0.82	1.42	1.22
Average	7.7	6.2	8.0	4.4	1.08	0.95	1.23	1.13

^b Obtained as $\frac{\mathbf{GSI}_{Ei+1}}{\mathbf{GSI}_{Ei}}$ $i = 0,1,2$.

Table S3. Time required, in simulated data, for the genomic selection index (L_{GSI}) and phenotypic selection index (L_{PSI}) to complete one selection cycle; correlation or accuracy ($\rho_{H,GSI}$) between the net genetic merit (H) and the genomic selection index (GSI) and the square root of the heritability of phenotypic selection index (h_{PSI}). These parameters were used to obtain the right-hand side ($\frac{\rho_{H,GSI}}{h_{PSI}} L_{PSI}$) of the inequality formula of Technow *et al.* (2013). The left-hand side of this inequality formula (L_{GSI}) was equal to 1.5 in all selection cycles.

Cycle	L_{GSI}	L_{PSI}	$\rho_{H,GSI}$	h_{PSI}	$\frac{\rho_{H,GSI}}{h_{PSI}} L_{PSI}$
1	1.5	4.0	0.73	0.92	3.17
2	1.5	4.0	0.78	0.89	3.50
3	1.5	4.0	0.83	0.88	3.77
4	1.5	4.0	0.74	0.87	3.40
5	1.5	4.0	0.71	0.87	3.30
6	1.5	4.0	0.73	0.84	3.49
7	1.5	4.0	0.80	0.85	3.80

Table S4. Time required in cycle 0 in four real data sets for the genomic selection index (L_{GSI}) and the phenotypic selection index (L_{PSI}) to complete the selection cycle; correlation or accuracy ($\rho_{H,GSI}$) between the net genetic merit (H) and the genomic selection index (GSI) and the square root of the heritability of phenotypic selection index (h_{PSI}). These parameters were used to obtain the right-hand side ($\frac{\rho_{H,GSI}}{h_{PSI}} L_{PSI}$) of the inequality formula of Technow *et al.* (2013). The left-hand side of this inequality (L_{GSI}) formula was equal to 1.5 in all selection cycles.

Real data set	L_{GSI}	L_{PSI}	$\rho_{H,GSI}$	h_{PSI}	$\frac{\rho_{H,GSI}}{h_{PSI}} L_{PSI}$
3	1.5	4.0	0.89	0.80	4.44
4	1.5	4.0	0.59	0.64	3.70
5	1.5	4.0	0.62	0.63	3.91
6	1.5	4.0	0.72	0.72	3.98