# Chapter 10
# Stochastic Simulation of Four Linear Phenotypic Selection Indices

**Fernando H. Toledo, José Crossa, and Juan Burgueño**

**Abstract** Stochastic simulation can contribute to a better understanding of the problem, and has already been successfully applied to evaluate other breeding scenarios. Despite all the theories developed in this book concerning different types of indices, including phenotypic data and/or data on molecular markers, no examples have been presented showing the long-term behavior of different indices. The objective of this chapter is to present some results and insights into the in silico (computer simulation) performance comparison of over 50 selection cycles of a recurrent and generic population breeding program with different selection indices, restricted and unrestricted. The selection indices included in this stochastic simulation were the linear phenotypic selection index (LPSI), the eigen selection index method (ESIM), the restrictive LPSI, and the restrictive ESIM.

## 10.1  Stochastic Simulation

Simulations were used to evaluate the accuracy, effectiveness, response to selection, and the decrease in the overall genetic variance in a recurrent selection scheme under the use of the Smith (1936) and Hazel (1943) index (or linear phenotypic selection index, LPSI, see Chap. 2 for details); the eigen selection index method (ESIM, see Chap. 7 for details); the Kempthorne and Nordskog (1959) restricted index (K&N or restricted phenotypic selection index, RLPSI, see Chap. 3 for details); and the restricted eigen selection index method (RESIM, see Chap. 3 for details). The different scenarios are described below and encompass variations in the nature of the genetic correlation between traits in addition to their expected heritabilities.

F. H. Toledo (✉) · J. Crossa · J. Burgueño
Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), Mexico, Mexico
e-mail: f.toledo@cgiar.org

### 10.1.1    Breeding Design

A total of 50 forward recurrent selection cycles of modern breeding were simulated, in which the breeder has the ability to select based on breeding value estimates of genetically correlated traits, and to apply the various above-mentioned selection indices. All simulated scenarios (described below) followed a common general breeding design. In each cycle, 350 full sib progenies ($S_1$) were generated taking 700 parents at random from the base population. From each progeny, 100 double-haploid lines were randomly derived (which shortened the cycle interval by five inbreeding generations). The simulated phenotypic values of the 35,000 resulting lines were then evaluated in simulated trials. The selection was made by means of progeny average performance. The selected progenies (top quarter) according to each index were then recombined by random mating a sample of the lines within the progeny to recover the population for the next cycle.

### 10.1.2    Simulating Quantitative Traits

Genetically correlated quantitative traits were simulated assuming a full pleiotropic model. This was carried out by randomly sampling genetic effects for all segregating sites from a multivariate normal distribution with zero mean and a previously stated variance–covariance. The genetic effects were in turn used to compute true breeding values (TBVs). An individual's phenotype was obtained by taking its TBV and adding a zero mean normally random term with variance consistent with the expected heritability ($h^2$) for the trait at which phenotyping occurred. The genetic variance in each cycle was calculated as the variance of the TBV of the individuals in that generation. However, it was expressed as relative values of the genetic variance in the initial cycle. The realized response to selection was also standardized in units of the genetic standard deviation in cycle 0. Cycle 0 was used as the base generation because it represents the available genetic variability, and also to observe, from the start, the genetic changes in future breeding generations.

An empirical genome was considered comprising a set of 10 linkage groups (chromosomes), each 200 cM in length, and 1000 uniformly distributed segregating sites. To represent the historical evolution and recent breeding efforts up to the present day in addition to incorporating a steady state of known linkage disequilibrium (LD) structure existing in crops, the starting populations (cycle 0) were taken after 200 generations of random mating within an effective population size of 1000 segregating for all loci in which the allele frequency was 0.5.

The in silico meiosis reflected the Mendelian laws of segregation for diploid species, by a count-location process that mimics the Haldane map function (Haldane 1919). Thus, homologous chromosomes are paired into bivalents and recombined through randomly positioned chiasmata. The number of chiasmata follows a Poisson distribution, where the λ parameter represents the chromosome length in Morgans and their positions are uniformly distributed, i.e., without interference between crossovers or any mutagenesis process.

### 10.1.3   Simulated Scenarios

Three traits were considered, one with low heritability (the first, $h^2 = 0.2$) and two with high heritability (the second and the third, $h^2 = 0.5$). The correlations between the first and second trait vary from positive ($\rho_G = 0.5$) to negative ($\rho_G = -0.5$). The third trait was always considered with segregation independent from the two others.

The selection process involved two unrestricted indices: the LPSI (see Chap. 2), which ranks the progenies based on the average merit of their lines considering equal economic weights for all traits, and the ESIM (see Chap. 7), where the progenies were ranked in terms of ESIM values. Regarding the restricted selection indices, the RLPSI (or K&N) was employed (see Chap. 3) with equal economic weights for the traits in addition to the RESIM (see Chap. 7). Because of the restrictions, two different situations were evaluated in the latter cases, i.e., where the restrictions were applied for each of the first and second traits separately.

Thus, all simulated scenarios encompass a three-way factorial: four selection procedures (the LPSI, the ESIM, the RLPSI or K&N, and the RESIM); two correlation scenarios, positive ($\rho_G = 0.5$) and negative correlations ($\rho_G = -0.5$) between the first and second trait; and two constraint situations, where the restrictions were applied separately for the first and second traits.

To simulate genetically correlated traits a full pleiotropic model was assumed. Gene effects were sampled from a multivariate normal distribution with zero mean and a previously stated variance–covariance matrix. In that sense it is possible to represent a quantitative and infinitesimal model. Each genes has its own effect varying according to a probabilistic density i.e., genes with positive and negative effects varying its effects sizes; alleles with large effects at lower frequency (major genes) and alleles with modest effects at higher frequency (minor genes).

### 10.1.4   Inferences

Results are presented as summaries of 100 Monte Carlo replicates for each scenario and include the response to selection, decreases in the genetic variance, selection accuracy, and observed heritabilities. The meiosis routine was implemented in C++, and compiled, linked, and through the facilities provided by the Rccp R package (Eddelbuettel 2013). All simulations were performed, analyzed, and summarized in R version 3.3.3 (R Development Core Team 2017).

## 10.2   Results

Overriding the results of the simulations regarding the four selection indices under the different trait genetic correlations and restrictions, scenarios are presented in terms of the consistency of the observed heritabilities of the traits; the response to

selection and changes in genetic variance for each trait; and the accuracy of the indices' selection.

First of all, the results show the stability of the Monte Carlo replicates in terms of possible deviations in the observed heritability from that expected, which in turn may affect further inferences (Table 10.1). The *type I* error ($\alpha$) of the *t test* comparing expected and observed heritabilities for all simulated scenarios did not show important and significant departures. Slight departures that may be due to Monte Carlo error ($P < 0.05$) were found, namely: for both high and low heritability traits of the LPSI at cycle 5 when they were negatively correlated; for the independent trait also with the LPSI at cycle 50, but, when the other traits are positively correlated; for the high heritable trait at the first and last cycles, both under positive correlation in the ESIM and RESIM indices respectively; and for the low heritability trait in both restricted indices (RLPSI and RESIM) in cycles 0 and 5 for respective and negative and positive correlations.

A complementary estimate of the power (*type II* error or $\beta$) of the tests was performed considering departures from the expected heritabilities of 1%. It was verified that the average power if the observed estimates was around 70%, which reinforces the appropriateness of the simulation findings.

### 10.2.1   Realized Genetic Gains

Figure 10.1 shows the average genetic gains (expressed as standard deviations from the mean of cycle 0) for cycles 0–50 for the traits (low and high heritabilities and the independent trait); the four selection indices (unrestricted: LPSI and ESIM and restricted: RLPSI and RESIM) when the correlations are positive and negative.

It is important to note that even after 50 recurrent cycles none of the scenarios has shown any indication that the selection plateau has been reached (Fig. 10.1). It is considered that even with the variation of the gains in the scenarios, there were increases in the merit of the target traits. Thus, the employment of selection indices is an effective way of achieving progress in long-term multi-trait selection.

As expected, the unrestricted selection indices have shown genetic gains higher than their restricted counterparts (Fig. 10.1). It must be highlighted that the restrictions proved their properties because when any trait was restricted, no gains were obtained for that trait (data not shown). The higher gains obtained with unrestricted indices is well known and justified in comparison with their restricted homologous because the net genetic merit is beneficiated by the gains in all traits, while, with gains constrained to zero in some traits, there are no indirect gains that may be highlighted especially because of positive correlations.

The independent trait has presented the higher gains in comparison with the other traits for all correlation and selection process scenarios. The higher gains, however, were for the RESIM followed by the RLPSI in both positive and negative correlations (Fig. 10.1e and f). These findings may be understood both under the nature of the trait (independent inheritance) and over the properties of the restricted indices.

**Table 10.1** Mean ($\mu$), standard deviation ($\sigma$), *type I* error ($\alpha$), and *type II*/power error ($\beta$) of the observed heritability in 100 Monte Carlo replicates for traits with low and high heritability ($h^2$) and independent at cycles 0, 5, and 50 of a simulated selection given four indices, the linear phenotypic selection index (LPSI), the ESIM, the restricted linear phenotypic selection index (RLPSI), and the RESIM, with positive and negative correlations between the traits low $h^2$ and high $h^2$

| Scenario/cycle | | Low $h^2$ | | Errors | | High $h^2$ | | Errors | | Independent | | Errors | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\mu$ | $\sigma$ | $\alpha$ | $\beta$ | $\mu$ | $\sigma$ | $\alpha$ | $\beta$ | $\mu$ | $\sigma$ | $\alpha$ | $\beta$ |
| **LPSI** | | | | | | | | | | | | | |
| Negative | 0 | 0.201 | 0.016 | 0.565 | 0.993 | 0.502 | 0.035 | 0.536 | 0.513 | 0.502 | 0.035 | 0.524 | 0.531 |
| | 5 | 0.204 | 0.016 | 0.016 | 0.993 | 0.493 | 0.031 | 0.028 | 0.612 | 0.504 | 0.034 | 0.191 | 0.540 |
| | 50 | 0.200 | 0.015 | 0.852 | 0.996 | 0.501 | 0.035 | 0.672 | 0.524 | 0.505 | 0.034 | 0.124 | 0.547 |
| Positive | 0 | 0.200 | 0.014 | 0.782 | 0.999 | 0.500 | 0.037 | 0.902 | 0.485 | 0.503 | 0.033 | 0.347 | 0.556 |
| | 5 | 0.200 | 0.016 | 0.886 | 0.995 | 0.499 | 0.032 | 0.758 | 0.601 | 0.502 | 0.033 | 0.561 | 0.562 |
| | 50 | 0.200 | 0.015 | 0.950 | 0.997 | 0.504 | 0.035 | 0.214 | 0.513 | 0.507 | 0.035 | 0.046 | 0.511 |
| **ESIM** | | | | | | | | | | | | | |
| Negative | 0 | 0.200 | 0.013 | 0.711 | 1.000 | 0.502 | 0.035 | 0.552 | 0.529 | 0.503 | 0.038 | 0.371 | 0.464 |
| | 5 | 0.201 | 0.015 | 0.677 | 0.997 | 0.497 | 0.033 | 0.349 | 0.559 | 0.502 | 0.031 | 0.445 | 0.611 |
| | 50 | 0.200 | 0.015 | 0.749 | 0.997 | 0.503 | 0.034 | 0.309 | 0.548 | 0.499 | 0.035 | 0.815 | 0.514 |
| Positive | 0 | 0.199 | 0.015 | 0.559 | 0.996 | 0.493 | 0.032 | 0.029 | 0.593 | 0.502 | 0.030 | 0.530 | 0.642 |
| | 5 | 0.200 | 0.014 | 0.892 | 0.999 | 0.498 | 0.032 | 0.490 | 0.589 | 0.505 | 0.037 | 0.222 | 0.472 |
| | 50 | 0.202 | 0.018 | 0.405 | 0.975 | 0.504 | 0.029 | 0.132 | 0.679 | 0.500 | 0.033 | 0.888 | 0.560 |
| **RLPSI** | | | | | | | | | | | | | |
| Negative | 0 | 0.198 | 0.016 | 0.178 | 0.994 | 0.497 | 0.033 | 0.405 | 0.557 | 0.503 | 0.032 | 0.438 | 0.587 |
| | 5 | 0.203 | 0.014 | 0.042 | 0.999 | 0.505 | 0.033 | 0.110 | 0.567 | 0.501 | 0.034 | 0.471 | 0.541 |
| | 50 | 0.202 | 0.015 | 0.173 | 0.996 | 0.501 | 0.032 | 0.742 | 0.602 | 0.500 | 0.033 | 0.397 | 0.572 |
| Positive | 0 | 0.201 | 0.014 | 0.295 | 0.999 | 0.504 | 0.033 | 0.239 | 0.570 | 0.502 | 0.034 | 0.130 | 0.539 |
| | 5 | 0.201 | 0.016 | 0.709 | 0.990 | 0.507 | 0.035 | 0.059 | 0.510 | 0.502 | 0.034 | 0.689 | 0.557 |
| | 50 | 0.201 | 0.013 | 0.331 | 1.000 | 0.504 | 0.032 | 0.195 | 0.591 | 0.507 | 0.035 | 0.061 | 0.533 |

(continued)

**Table 10.1** (continued)

| Scenario/cycle | | Low $h^2$ | | Errors | | High $h^2$ | | Errors | | Independent | | Errors | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\mu$ | $\sigma$ | $\alpha$ | $\beta$ | $\mu$ | $\sigma$ | $\alpha$ | $\beta$ | $\mu$ | $\sigma$ | $\alpha$ | $\beta$ |
| RESIM | | | | | | | | | | | | | |
| Negative | 0 | 0.201 | 0.017 | 0.504 | 0.987 | 0.506 | 0.034 | 0.075 | 0.545 | 0.502 | 0.033 | 0.646 | 0.567 |
| | 5 | 0.202 | 0.016 | 0.230 | 0.993 | 0.506 | 0.032 | 0.090 | 0.585 | 0.500 | 0.031 | 0.757 | 0.618 |
| | 50 | 0.199 | 0.015 | 0.490 | 0.996 | 0.507 | 0.034 | 0.051 | 0.549 | 0.501 | 0.032 | 0.516 | 0.604 |
| Positive | 0 | 0.203 | 0.014 | 0.018 | 0.999 | 0.506 | 0.035 | 0.075 | 0.510 | 0.505 | 0.033 | 0.172 | 0.578 |
| | 5 | 0.202 | 0.014 | 0.171 | 0.998 | 0.502 | 0.034 | 0.489 | 0.555 | 0.501 | 0.031 | 0.734 | 0.626 |
| | 50 | 0.200 | 0.017 | 0.911 | 0.988 | 0.508 | 0.036 | 0.034 | 0.495 | 0.501 | 0.032 | 0.694 | 0.601 |
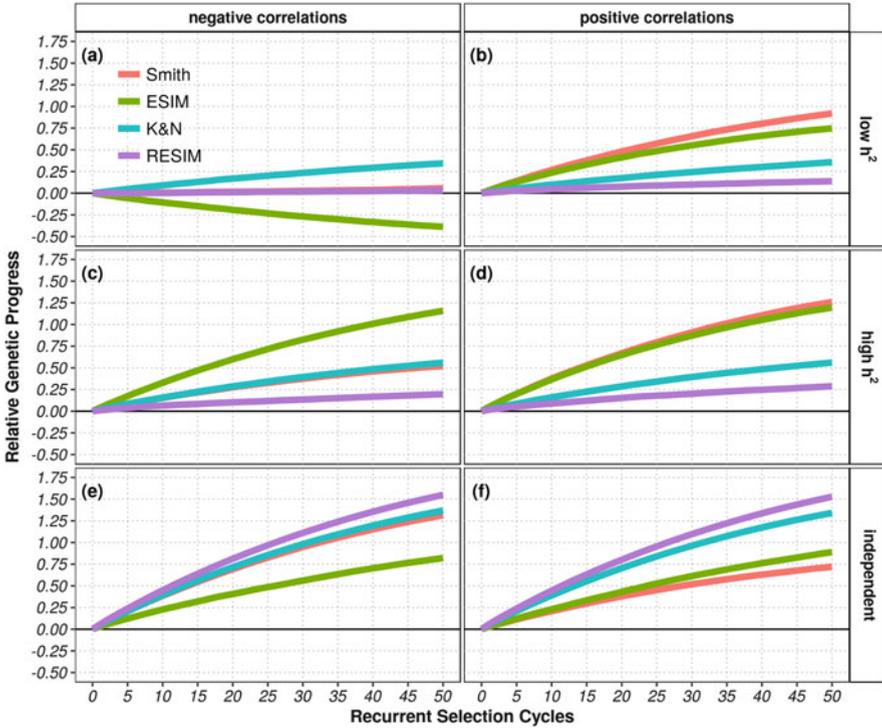
**Fig. 10.1** Average genetic gains in 100 Monte Carlo replicates for traits with low and high heritability ($h^2$ 0.2 and 0.5) and independent along cycles 0–50 of a simulated selection given four indices, the linear phenotypic selection index (LPSI), the ESIM, the restricted linear phenotypic selection index (RLPSI), and the RESIM with positive (0.5) and negative (−0.5) correlations between the traits low $h^2$ and high $h^2$. (**a**) Gains for the trait with low heritability when it is negatively correlated with the high heritability trait. (**b**) Gains for the trait with low heritability when it is positively correlated with the high heritability trait. (**c**) Gains for the trait with high heritability when it is negatively correlated with the low heritability trait. (**d**) Gains for the trait with high heritability when it is negatively correlated with the low heritability trait. (**e**) Gains for the independent trait when the other traits are negatively correlated. (**f**) Gains for the independent trait when the other traits are positively correlated

As the third trait becomes independent from the others, there are no indirect effects owing to the constraints in the gains of the other traits. With regard to the technical features of the RESIM, it must be emphasized that because of the eigen decomposition, the largest eigenvector obtains higher weight from the most variable trait and consequently ends in distinct gains, which in this case is the independent trait.

The Smith (or LPSI) and ESIM produce similar genetic gains for highly heritable traits when the genetic correlations are positive (Fig. 10.1d). The ESIM is simply another way of obtaining the LPSI based on the eigen decomposition theory, which avoids the assignment of economic weights. Thus, the results prove that the same results may be found with both indices. However, the ESIM is the preferred index

because of its advantages over the LPSI: no subjective decision for selecting economic weights, and better statistical sampling properties.

When the traits are negatively correlated, the trait with greater heritability has shown important realized genetic gains based on the ESIM and similar gains for the LPSI and its restricted analogous, i.e., the RLPSI (Fig. 10.1a and c). In addition, when traits are negatively correlated, restricting the traits with low heritability is an alternative, to ensure similar progress to the use of unrestricted indices for highly heritable traits. On the contrary, it is also interesting to note that the ESIM has the worst performance when the traits are negatively correlated for trait with lower heritability (Fig. 10.1a).

On the other hand, as already pointed out, the ESIM performance surpasses all the others with regard to the highly heritable trait (Fig. 10.1c and d). The reason for this is similar to the above-mentioned regarding the properties of the eigen decomposition. When the first trait is negatively correlated with the second one, heavier weight is given to the trait with higher heritability than to the trait with low heritability. However, when the traits are positively correlated, synergic and indirect effects increase both traits, one positively affecting the other.

When the traits are positively correlated but with low heritability, the LPSI and the ESIM have similar realized genetic gains until cycle 25; after this selection cycle, the LPSI is superior to the ESIM (Fig. 10.1b). In this case, the two restrictive indices, the RLPSI and the RESIM, are given lower realized genetic gains than the LPSI and the ESIM (Fig. 10.1b). Finally, considering the third trait (the independent one), the RESIM provides the greater realized genetic gains (Fig. 10.1e and f).

## 10.2.2   Genetic Variances

In Fig. 10.2, the average relative decreases in the genetic variances along the 50 cycles of selection for the three traits (with low and high heritability traits in addition to the independent trait) under the selection system given by the four selection indices, restricted (the RLPSI and the RESIM) and unrestricted (the LPSI and the ESIM), both with negative and positive correlations between the first and second traits.

As a general result, it is clear that after selection there were decreases in the genetic variance along the recurrent cycles (Fig. 10.2). From the most conservative decrease (around 40% in Fig. 10.2a and b) to the sharp decrease (close to 10% in Fig. 10.2e and f) and in contrast to the trends in genetic gains, it is possible to conceive that the genetic variability was not yet exhausted by selection. This observation endorses what was said regarding the effectiveness of the selection indices as a criterion for long-term multi-trait selection.

As expected, the restricted indices are more conservative, maintaining greater genetic variance (Fig. 10.2). Their feature is to prevent the restricted trait from changing its genetic merit. Thus, they tend to keep its genetic variance unchanged,
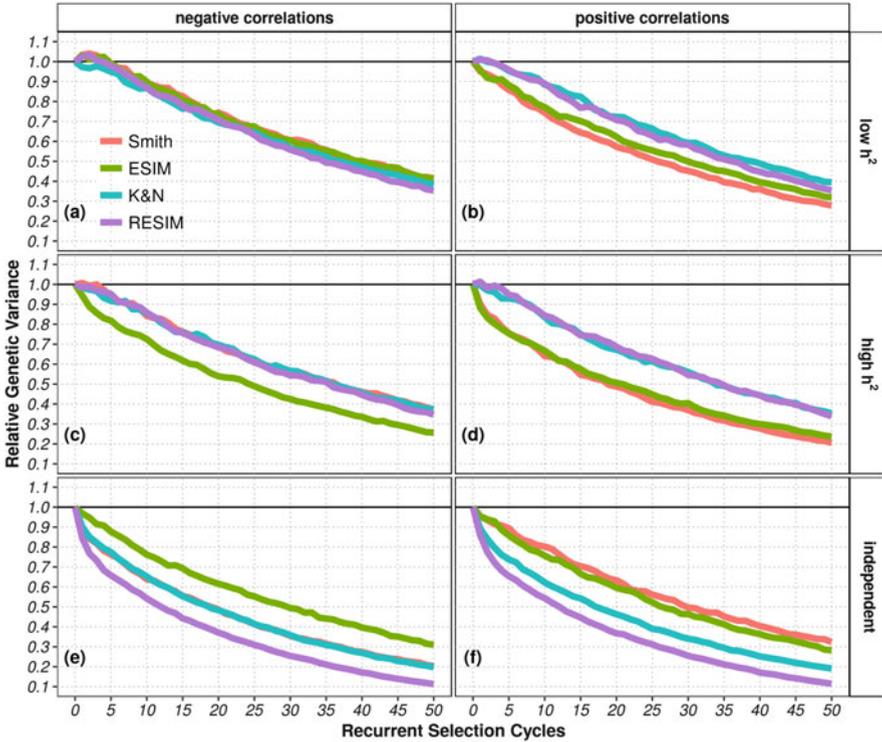
**Fig. 10.2** Average genetic variances in 100 Monte Carlo replicates for traits with low and high heritability ($h^2$ 0.2 and 0.5) and independent along cycles 0–50 of a simulated selection given four selection indices, the LPSI, the ESIM, the RLPSI, and the RESIM, with positive (0.5) and negative (−0.5) correlations between the traits low $h^2$ and high $h^2$. (**a**) Genetic variance of the low heritability trait when it is negatively correlated with the high heritability trait. (**b**) Genetic variance of the low heritability trait when it is positively correlated with the high heritability trait. (**c**) Genetic variance of the high heritability trait when it is negatively correlated with the low heritability trait. (**d**) Genetic variance of the high heritability trait when it is negatively correlated with the low heritability trait. (**e**) Genetic variance of the independent trait when the other traits are negatively correlated. (**f**) Genetic variance of the independent trait when the other traits are positively correlated

which is reflected in the lower decreases in the genetic variance, even under the indirect effects of the other traits.

It should be noted that there was a slight increase in variance in the short term (up to cycle 3) for the trait with lower heritability when negatively correlated with the highly heritable one (Fig. 10.2a and b). This is an outcome of the changes in allele frequencies of the first trait due to the indirect effects of the second trait and/or the release of genetic disequilibrium owing to the assortative mating of the individuals given higher weights regarding the second trait (highly heritable).

Reflecting the findings regarding the genetic gains (Fig. 10.1), the trait with strong decreases in genetic variance on average was the one in which the response

to selection was more pronounced, i.e., the independent trait (Fig. 10.2e and f). This trait has shown stronger decreases over the selection through the ESIM index in both positive and negative correlation scenarios. As mentioned before, as the third trait is independent of the others, a greater response to selection was achieved in that trait and consequently strong changes in allele frequencies, which drove the decreases in genetic variance.

When the heritability is high, it is easy to differentiate the trends in the decrease in the genetic variance between restricted and unrestricted indices (Fig. 10.2c). It is more evident, especially when the traits are positively correlated (Fig. 10.2d). Thus, the ESIM has the highest decreases followed by the LPSI. Nevertheless, for the traits with low heritability, the decreases in genetic variance are indistinguishable between the indices, showing that the effectiveness of the response to selection is a function of the heritability (Fig. 10.2a and b).

### 10.2.3   Selection Accuracy

The accuracy of the selection was measured as the square root of the correlation between the net genetic merit and the estimated linear function of each index. Figure 10.3 shows the absolute accuracies (left axis) and relative values in relation to the mean accuracy of the first cycle (right axis) for all indices in both negative (Fig. 10.3a) and positive (Fig. 10.3b) correlation scenarios.

In all cases, a reduction in the selection precision of all the indices was observed. The effect of selection is the improvement in the genetic merit of the traits by means of changes in allele frequencies that also affect/decrease the genetic variance. However, as a side effect, the selection becomes harder and has lower precision.

The LPSI has shown greater accuracy in comparison with the other indices in any situation (Fig. 10.3a and b). Its main feature is precisely maximizing the correlation between the net genetic merit and the linear combination of the trait. It may be
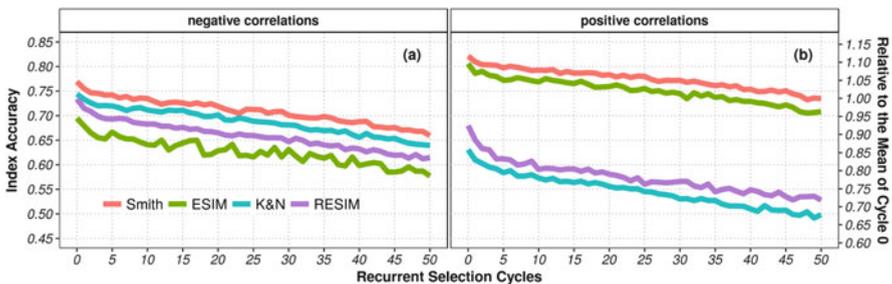


**Fig. 10.3** Average absolute and relative accuracy of selection in 100 Monte Carlo replicates for traits with low and high heritability ($h^2$) and independent along cycles 0–50 of a simulated selection given four selection indices, the LPSI, the ESIM, the RLPSI, and the RESIM with positive and negative correlations between the traits low $h^2$ and high $h^2$

argued that the ESIM also does that; however, only when the phenotypic and genotypic variances and covariances are known are they the best linear predictors. Thus, according to what was found, it is possible to note that the ESIM was more affected by the sampling properties when estimating matrices of variance and covariance (Fig. 10.3a).

For the scenario with positive correlations, the differences between the two types of indices, the restricted ones and the unrestricted ones, were clear, as the unrestricted indices have shown greater selection accuracy (Fig. 10.3b). This reflects the fact that the restricted index constrains the gains by means of restrictions in the correlation between the net genetic merit and the linear combination of the traits.

# References

Eddelbuettel D (2013) Seamless R and C++ Integration with Rcpp. Springer, New York
Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. J Genet 8:299–309
Hazel IN (1943) The genetics basis for constructing selection indexes. Genetics 28:476–490
Kempthorne O, Nordskog AW (1959) Restricted selection indices. Biometrics 15:10–19
R Core Team (2017) R: A language and environment for statistical computing
Smith HF (1936) A discriminant function for plant selection. Ann Eugenics 7:240–250