

Supplemental Methods

Principle and Mechanism of the Sparrow Search Algorithm

In genomic prediction, each individual is represented by a high-dimensional SNP genotype vector encoded as 0/1/2, and the full population forms an $n \times d$ genotype matrix. These SNP features define the high-dimensional input space on which machine-learning models are trained, and the same dimensionality naturally corresponds to the D-dimensional hyperparameter search space explored by SSA. Suppose that there are N sparrows in the D-dimensional search space and the position of the i th sparrow is:

$$X_i = X_{i1}, X_{i2}, X_{i3}, \dots, X_{id}, i = 1, 2, 3, \dots, n \quad (1)$$

In this context, the “position of the i th sparrow” is interpreted as one candidate hyperparameter configuration used to train genomic prediction models on the SNP genotype matrix. The location of the population X comprises n sparrows, as follows:

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,d} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n,1} & X_{n,2} & \dots & X_{n,d} \end{bmatrix} \quad (2)$$

where d denotes the dimension of the demand solution problem and $X_{i,d}$ represents the i -th sparrow. The fitness value F_x indicates the energy reserve, and finders with higher energy reserves are given preference. The fitness value F_x of all sparrow fitness values can be expressed as:

$$F_x = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_i \\ \vdots \\ f_n \end{bmatrix} = \begin{bmatrix} f(X_{1,1}, X_{1,2}, \dots, X_{1,d}) \\ f(X_{2,1}, X_{2,2}, \dots, X_{2,d}) \\ \vdots \\ f(X_{i,1}, X_{i,2}, \dots, X_{i,d}) \\ \vdots \\ f(X_{n,1}, X_{n,2}, \dots, X_{n,d}) \end{bmatrix} \quad (3)$$

For genomic prediction, the fitness function was defined as the mean squared error (MSE) between the observed phenotype vector y and the predicted values from the genotype matrix X . In SSA, after each iteration, the location update of the discoverer was described as follows:

$$x_{i,j}^{t+1} = \begin{cases} x_{i,j}^t \cdot \exp\left(\frac{-i}{\alpha \cdot iter_{max}}\right) & R_2 < ST \\ x_{i,j}^t + Q \cdot L & R_2 \geq ST \end{cases} \quad (4)$$

$x_{i,j}^t$ denotes the position of the j -th dimension of the i -th individual in the t -th generation of the population, t denotes the current number of iterations, and $iter_{max}$ is a constant representing the maximum number of iterations.

The parameter α is a uniform random number in the range $[0,1]$, that influences the search step size, whereas Q follows a standard normal distribution, introducing randomness to the search process. Matrix L is a $1 \times D$ -dimensional matrix with all elements equal to 1, ensuring a consistent adjustment across all dimensions.

In addition, R^2 is a uniform random number in $[0,1]$ that represents the alarm threshold that determines whether a sparrow enters an alert state. The safety threshold ST , also known as the warning value, was set to $[0.5, 1.0]$ to regulate the sparrow movement strategy, preventing premature convergence and enhancing global search capability.