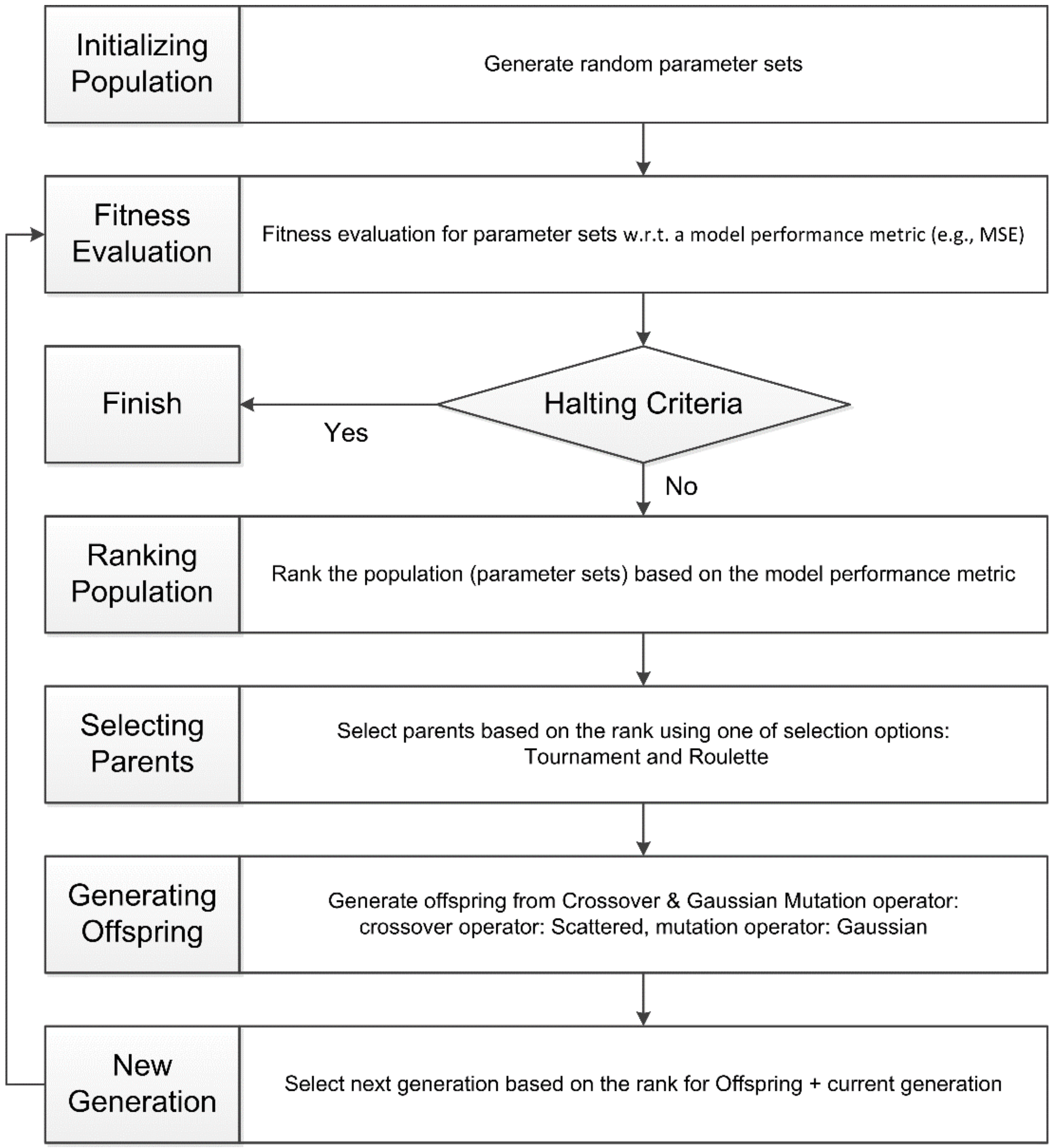
Note for Genetic Algorithm

- Genetic Algorithm diagram



- Genetic Algorithm Terminology

    - **Fitness functions**: The *fitness function* is the function you want to optimize. For standard optimization algorithm, this is known as the objective function.
    - **Individuals**: An *individual* is any point to which you can apply the fitness function. In the application of GA to calibrating hydrologic model parameters, a parameter set is an individual. The value of the fitness function for an individual is its score.
    - **Populations and Generations**: A *population* is an array of individuals (parameter sets). For example, if the size of the population is 100 and the number of variables (parameters) in the fitness function is 15, you represent the population by a 100-by-15 matrix.
      At each iteration, the genetic algorithm performs a series of computations on the current population to produce a new population. Each successive population is called a new *generation*.
    - **Parents and Children**: To create the next generation, the genetic algorithm selects certain individuals in the current population, called *parents*, and uses them to create individuals in the next generation, called *children*. Typically, the algorithm is more likely to select parents that have better fitness values.

- Stopping conditions for the algorithm

    - **Generation limit**: The algorithm stops when the number of generations reaches the value of **Generation limit**.
    - **Stall generations**: The algorithm stops when the average relative change in the fitness function value over **Stall generation**s is less than **Function tolerance**.
    - **Function Tolerance**: The algorithm runs until the average relative change in the fitness function value over Stall generations is less than Function tolerance.

- Reproduction options

    - **Population size**: Population size specifies how many individuals there are in each generation. With a large population size, the genetic algorithm searches the solution space more thoroughly, thereby reducing the chance that the algorithm returns a local minimum that is not a global minimum. However, a large population size also causes the algorithm to run more slowly.
    - **Elite count**: The number of individuals with the best fitness values in the current generation that are guaranteed to survive to the next generation. These individuals are called *elite children*.
    - **Crossover fraction**: The fraction of individuals in the next generation, other than elite children, that are created by crossover.

- Mutation and Crossover

  The genetic algorithm uses the individuals in the current generation to create the children that make up the next generation. Besides elite children, the algorithm creates

  − **Crossover children** by selecting genes (parameters) from a pair of individuals in the current generation and combines them to form a child.
  − **Mutation children** by applying random changes to a single individual in the current generation to create a child.

  Both processes are essential to the genetic algorithm. Crossover enables the algorithm to extract the best genes (parameters) from different individuals and recombine them into potentially superior children. Mutation adds to the diversity of a population and thereby increases the likelihood that the algorithm will generate individuals with better fitness values.

- Fitness scaling options

  Fitness scaling coverts the raw fitness cores that are returned by the fitness function to values in a range that is suitable for the selection function. **Scaling function** specifies the function that performs the scaling. The options are

  − **Rank**: Rank scales the raw scores based on the rank of each individual instead of its score. The rank of an individual is its position in the sorted scores. An individual with rank r has scaled score proportional to $1/\sqrt{r}$. So the scaled score of the most fit individual is proportional to 1, the scaled score of the next most fit is proportional to $1/\sqrt{2}$, and so on. Rank fitness scaling removes the effect of the spread of the raw scores. The square root makes poorly ranked individuals more nearly equal in score.

- Selection options

  Selection options specify how the genetic algorithm chooses parents for the next generation. **Selection function** specifies the function the algorithm uses to perform the selection. The options are

  − **Roulette**: Roulette selection chooses parents by simulating a roulette wheel, in which the area of the section of the wheel corresponding to an individual is proportional to the individual's expectation. The algorithm uses a random number to select one of the sections with a probability equal to its area.
  − **Tournament**: Tournament selection chooses each parent by choosing Tournament size players at random and then choosing the best individual out of that set to be a parent. Tournament size must be at least 2. The default value of Tournament size is 4.

- Crossover options

  Crossover options specify how the genetic algorithm combines two individuals, or parents, to form a crossover child for the next generation. **Crossover function** specifies the function that performs the crossover. The options are

  − **Scattered**: This is a crossover function for problems without linear constraints. This creates a random binary vector and selects the genes (model parameters) where the vector is a 1 from the first parent, and the genes where the vector is a 0 from the second parent, and combines the genes to form the child.

- Mutation options

  Mutation options specify how the genetic algorithm makes small random changes in the individuals in the population to create mutation children. Mutation provides genetic diversity and enables the genetic algorithm to search a broader space. **Mutation function** specifies the function that performs the mutation. The options are

  − **Gaussian**: Gaussian adds a random number taken from a Gaussian distribution with mean 0 to each entry of the parent vector. The standard deviation of this distribution is determined by the Scale (determining the standard deviation of the first generation), Shrink (controlling how the standard deviation shrinks as generations go by), and initial range setting. The default value of both Scale and Shrink is 1.