

1) Empirical data are collected, which has unknown parameters (θ) but known Summary Statistics (S_{obs}) :

$$[\theta_1, \theta_2 \dots \theta_n, S_1(obs), S_2(obs) \dots S_n(obs)]$$

2) Simulation data are generated, for multiple models of population histories which have known parameters (θ) and known Summary Statistics (S) :

$$\begin{aligned} \text{Model 1:} & \quad [\theta_1, \theta_2 \dots \theta_n, S_1, S_2 \dots S_n] \\ \text{Model 2:} & \quad [\theta_1, \theta_2 \dots \theta_n, S_1, S_2 \dots S_n] \\ \text{Model 3:} & \quad [\theta_1, \theta_2 \dots \theta_n, S_1, S_2 \dots S_n] \\ & \quad \vdots \\ \text{Model m:} & \quad [\theta_1, \theta_2 \dots \theta_n, S_1, S_2 \dots S_n] \end{aligned}$$

3) Standardized Sum of Squares is used to calculate the Euclidean distance between summary statistics of empirical data and simulated data:

$$\sqrt{\sum \left[\frac{S_n - S_n(obs)}{SD S_n} \right]^2}$$

4) Simulated models are ranked based on Euclidean distance. Epsilon (ϵ) is determined as the closest fraction (0.0005%) of models:

$$\epsilon \frac{\begin{aligned} \text{Model x:} & \quad [\theta_1, \theta_2 \dots \theta_n, S_1, S_2 \dots S_n] \\ \text{Model y:} & \quad [\theta_1, \theta_2 \dots \theta_n, S_1, S_2 \dots S_n] \\ \text{Model z:} & \quad [\theta_1, \theta_2 \dots \theta_n, S_1, S_2 \dots S_n] \\ & \quad \vdots \\ \text{Model m:} & \quad [\theta_1, \theta_2 \dots \theta_n, S_1, S_2 \dots S_n] \end{aligned}}{\text{---}}$$

5) The parameters (θ_m) of the top models are summarized into a distribution, which represents the approximate parameters of the empirical data (θ_{obs}). From this distribution, statistical estimators can be generated, such as credibility intervals, and the mean median or mode to represent point estimates.

