Seasonal Inhomogeneous Arrival Process Search and Evaluation

Paul Gibby

Kimberly Holmgren

Joseph R. Zipkin

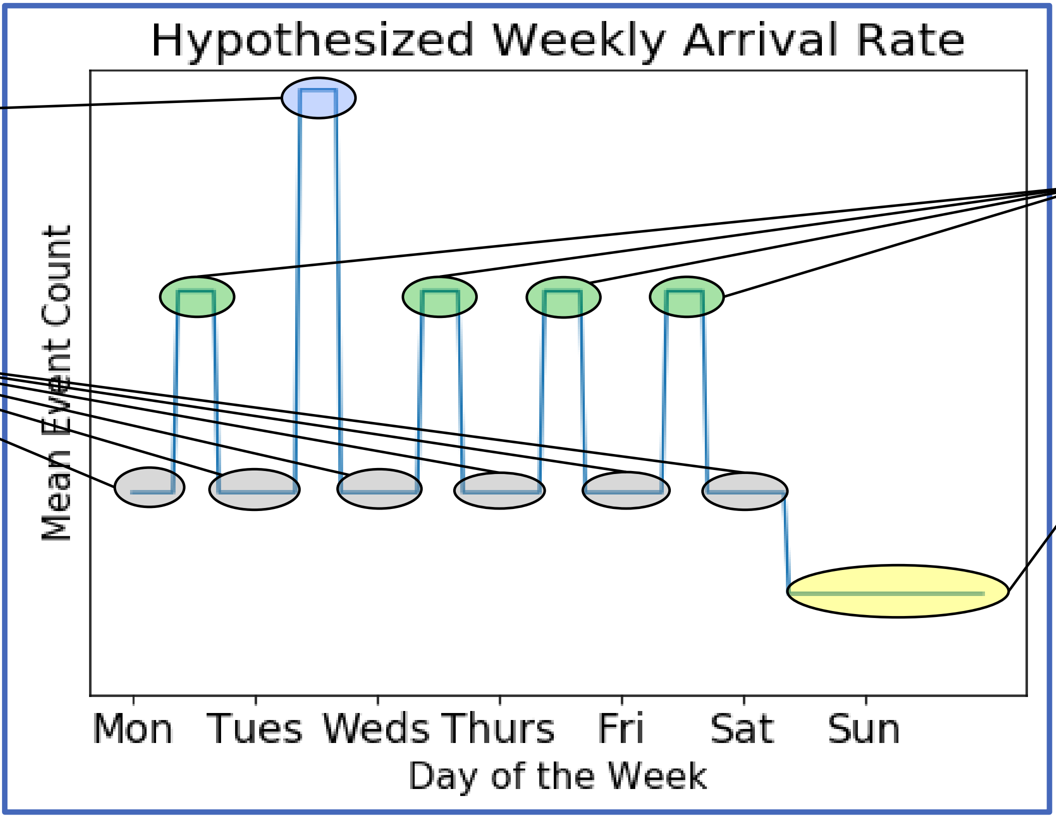
LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

This material is based upon work supported under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. Air Force.

Model Hypothesis: Cybersecurity Domain

Higher expected arrival rate on "Patch Tuesdays"

Rates overnight on weekdays lower than during the day but higher than over the weekend



Most attacks expected to occur during work hours – weekdays may be part of the same "regime"

Fewer attacks during weekends

Key Features

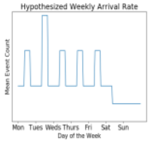
- Nonhomogeneous piecewise model
- Nonconsecutive intervals
- Repetitive
- Sparse data

Model Vocabulary

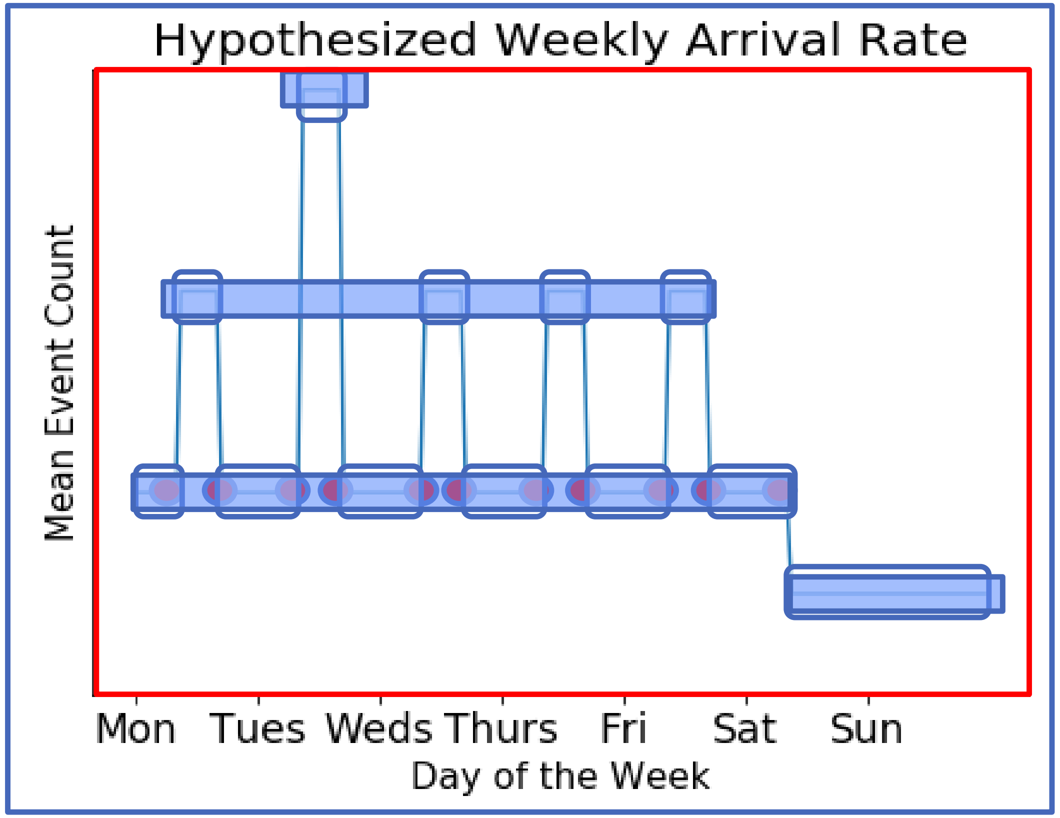
Breakpoints

Intervals

Regimes



Interval Division

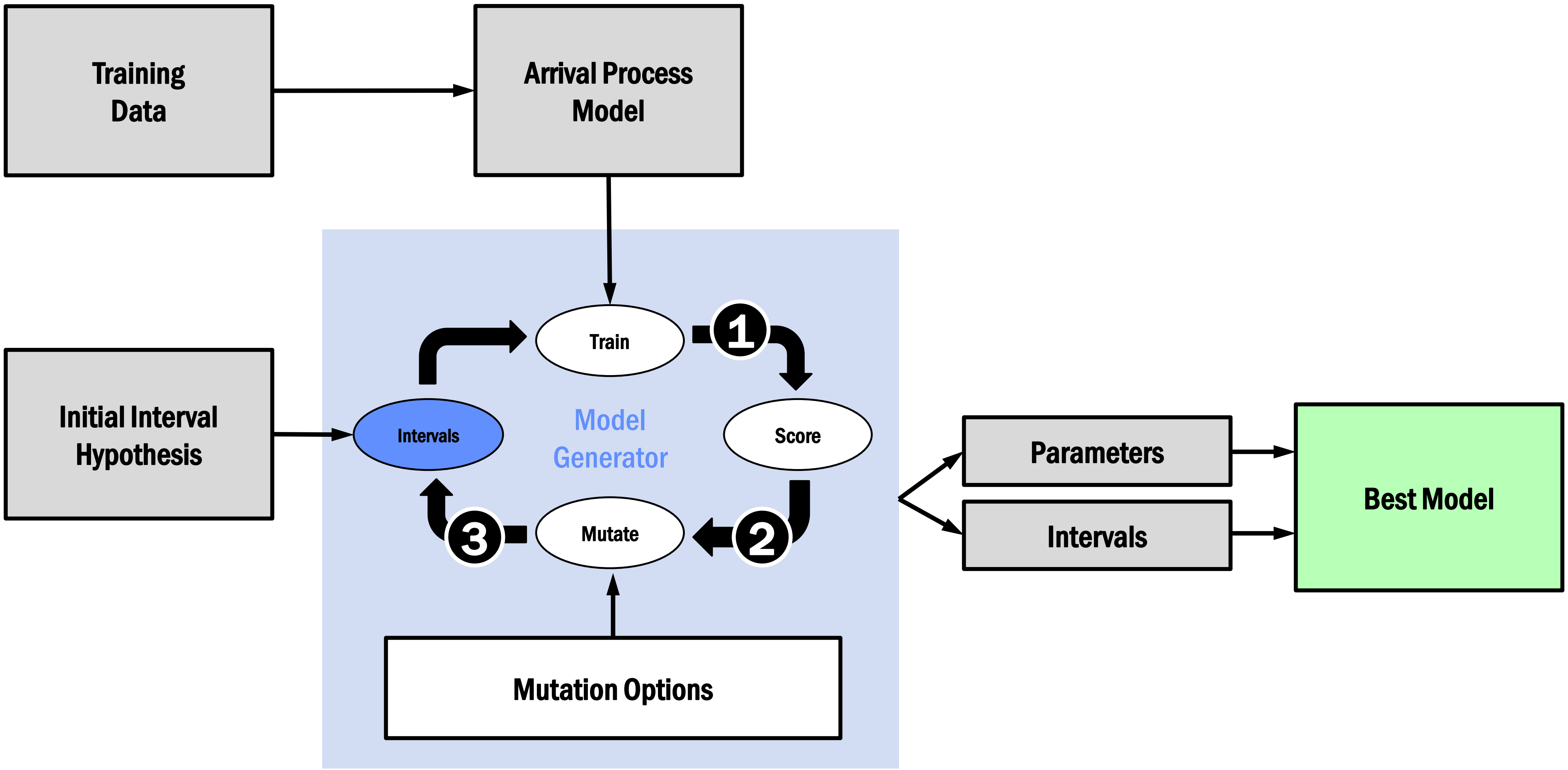


Questions

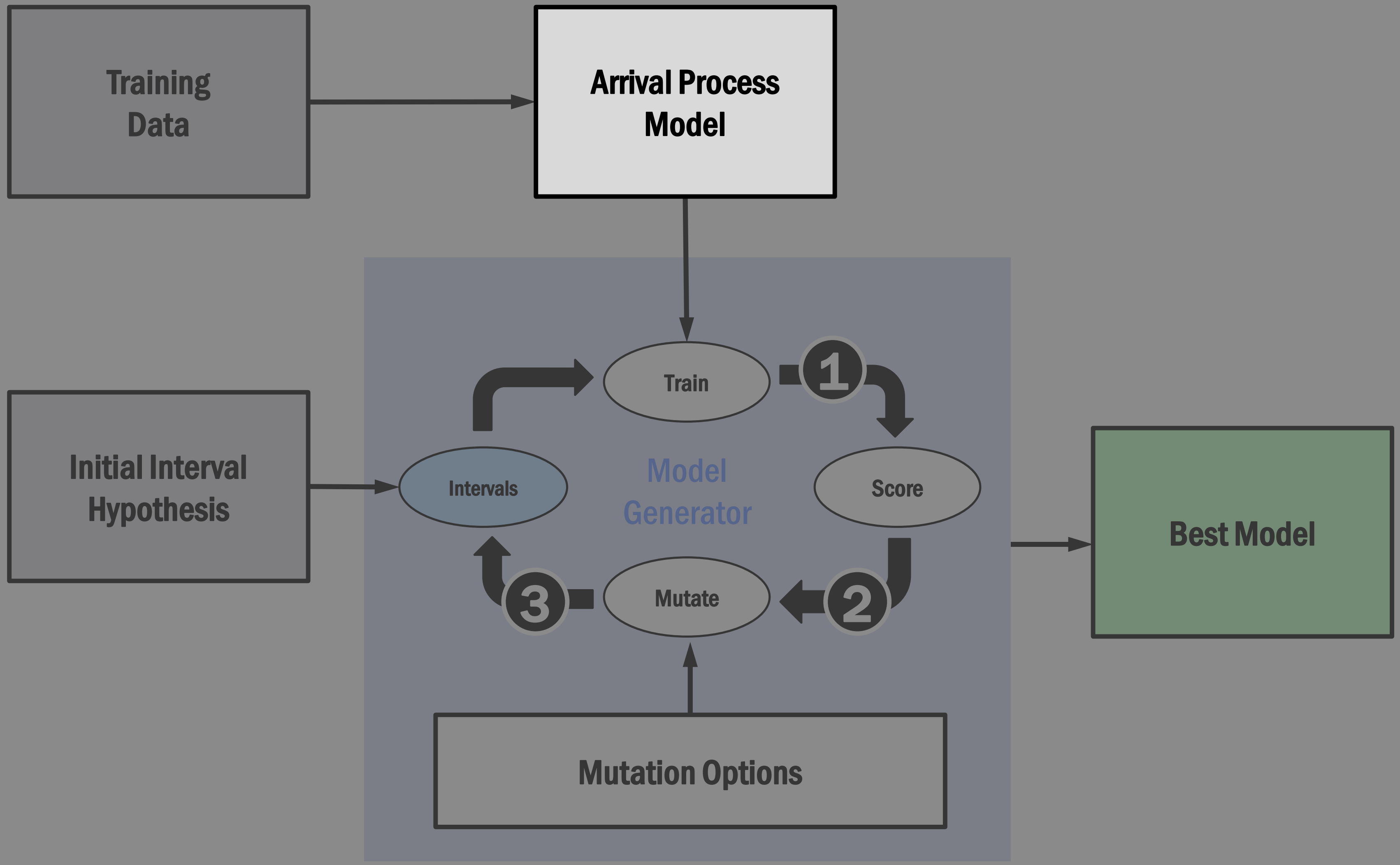
- Which model type fits best?
- Where are the breakpoints?
- How should intervals be assigned to regimes?
- What are the correct parameters for each regime?

SINAPSE Algorithm

Seasonal Inhomogeneous Arrival Process Search and Evaluation

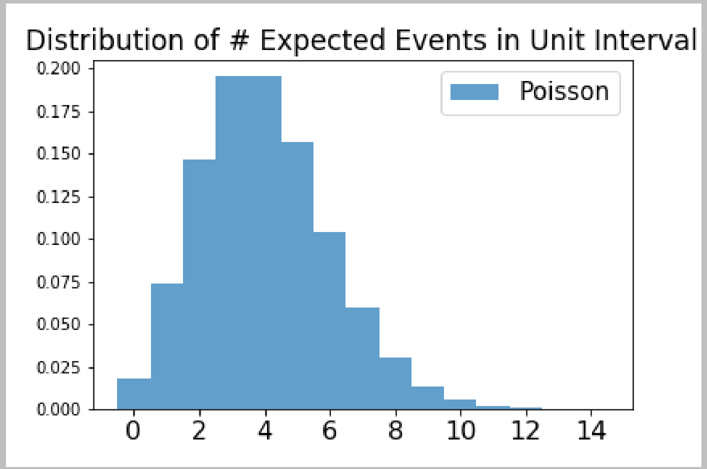


SINAPSE Algorithm



Arrival Process Models

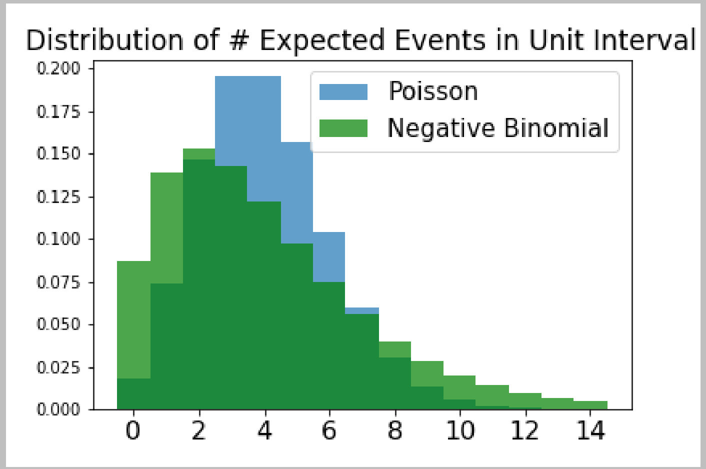
Poisson



$$P(N = n; \lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$$

- Models occurrence of events over time
- Time between events is independent, exponentially distributed

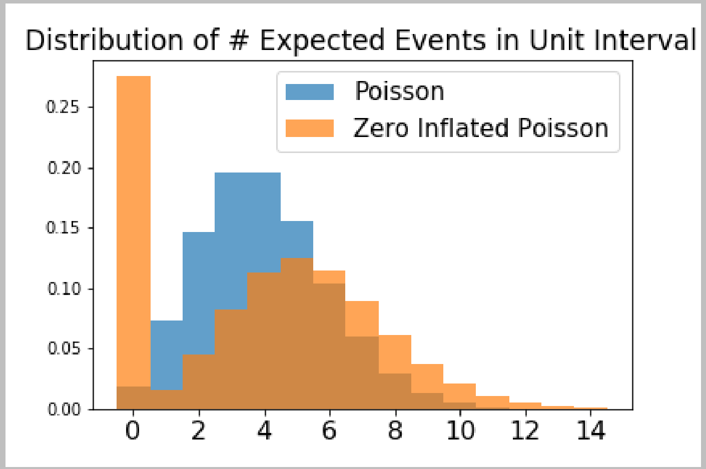
Negative Binomial



$$P(N = n; r, p) = \binom{n + r - 1}{n} (1 - p)^r p^n$$

- Variance may be greater than the mean
- More parameters, greater risk of overfitting

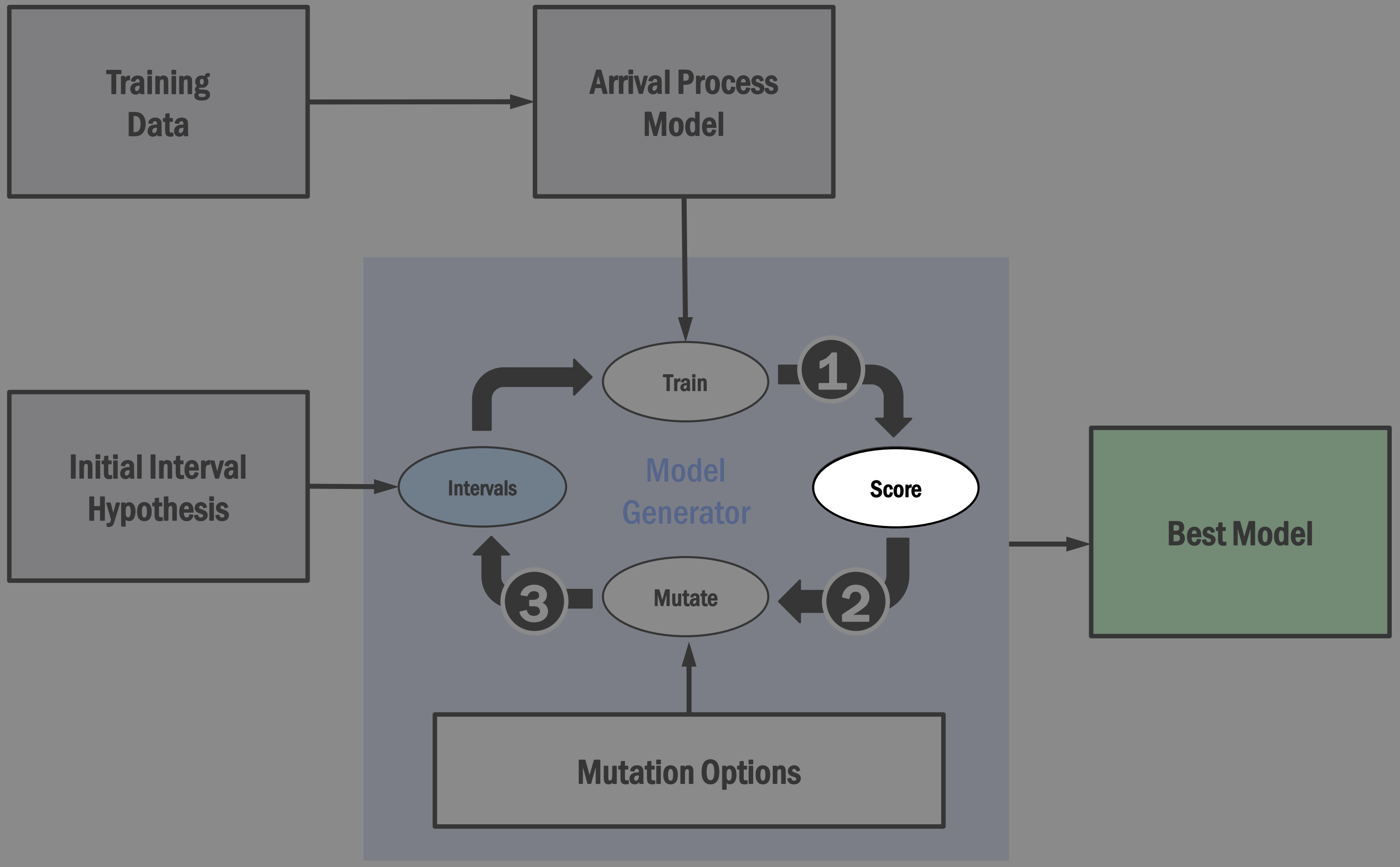
Zero-Inflated Poisson



$$P(N = n; \lambda, \pi) = \begin{cases} \pi + (1 - \pi) e^{-\lambda} & n = 0 \\ (1 - \pi) \frac{\lambda^n e^{-\lambda}}{n!} & otherwise \end{cases}$$

- Poisson with excess zeros from outside process

SINAPSE Algorithm



Measure of Fit

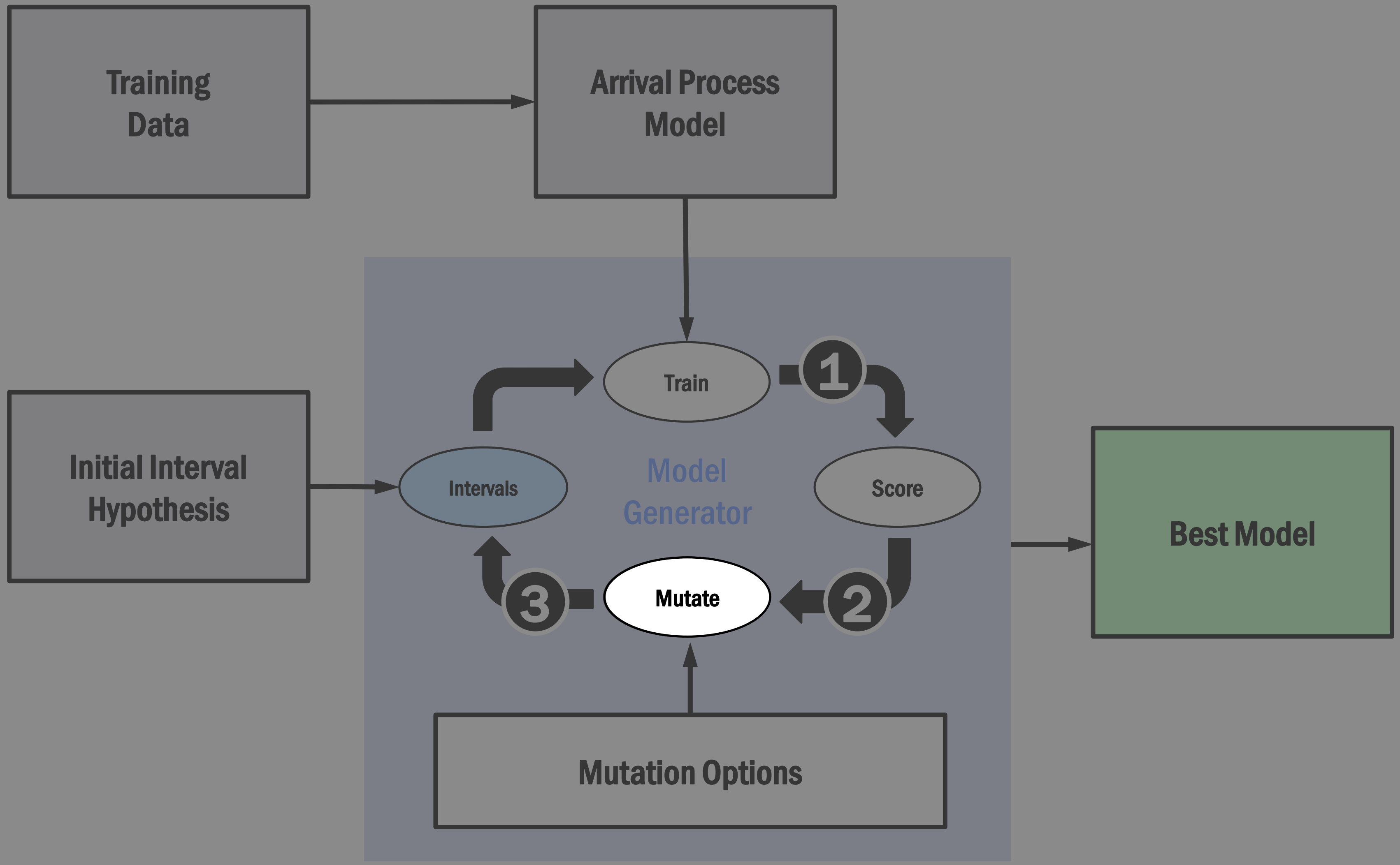
$$AICc = 2\boxed{P} - 2\log L(\theta|X) + \frac{2\boxed{P}(\boxed{P}+1)}{(n-\boxed{P}-1)}$$

Formula for Penalty Parameter P

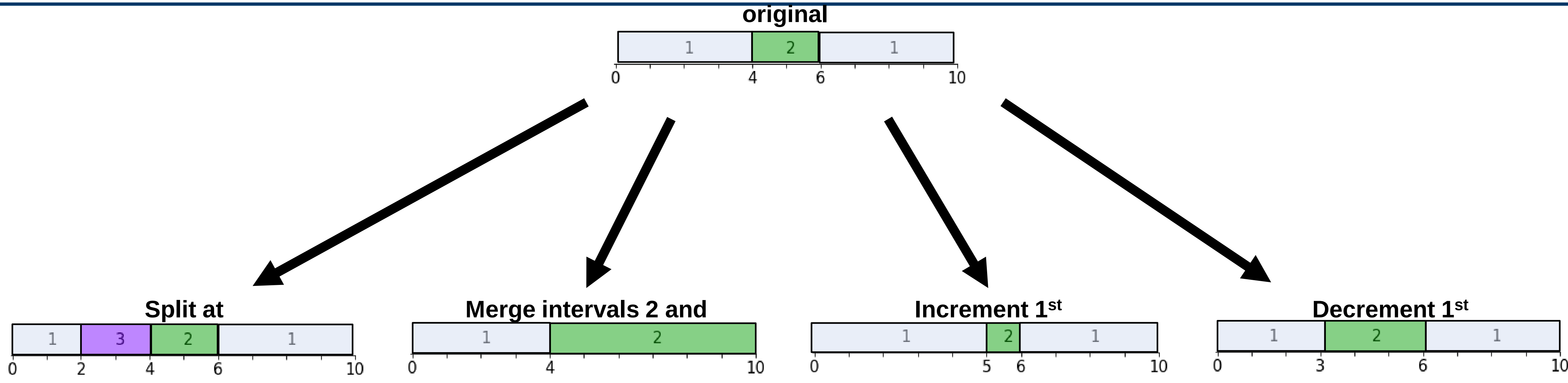
P = # Intervals + (# Regimes) * r

Arrival Process Model	Parameter Count per Regime (r)
Poisson	1
Negative Binomial	2
Zero Inflated Poisson	2

SINAPSE Algorithm



Mutation Options

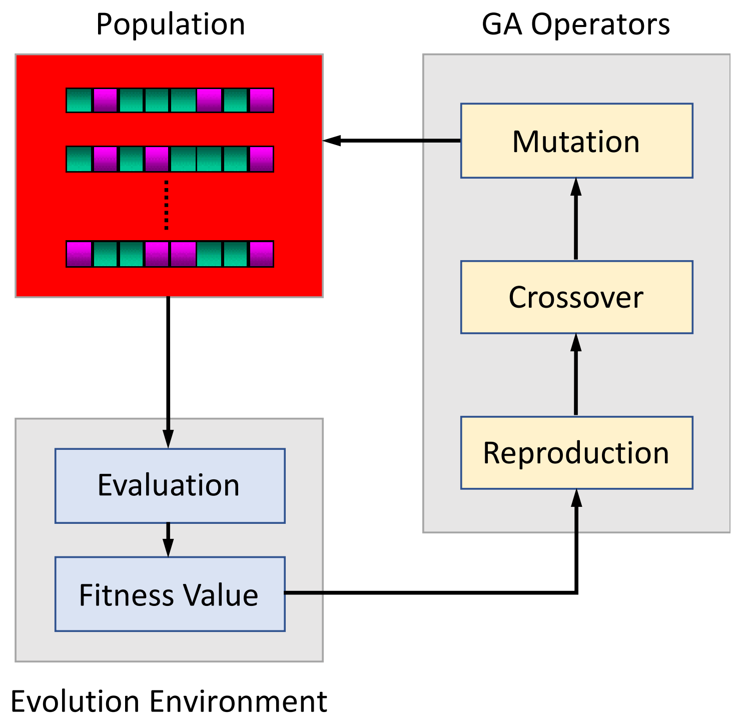


Interval Mutations

1. Randomly split an interval by adding a new breakpoint, creating a new regime
2. Randomly merge two intervals
3. Randomly increment the location of a breakpoint
4. Randomly decrement the location of a breakpoint

Genetic Algorithm

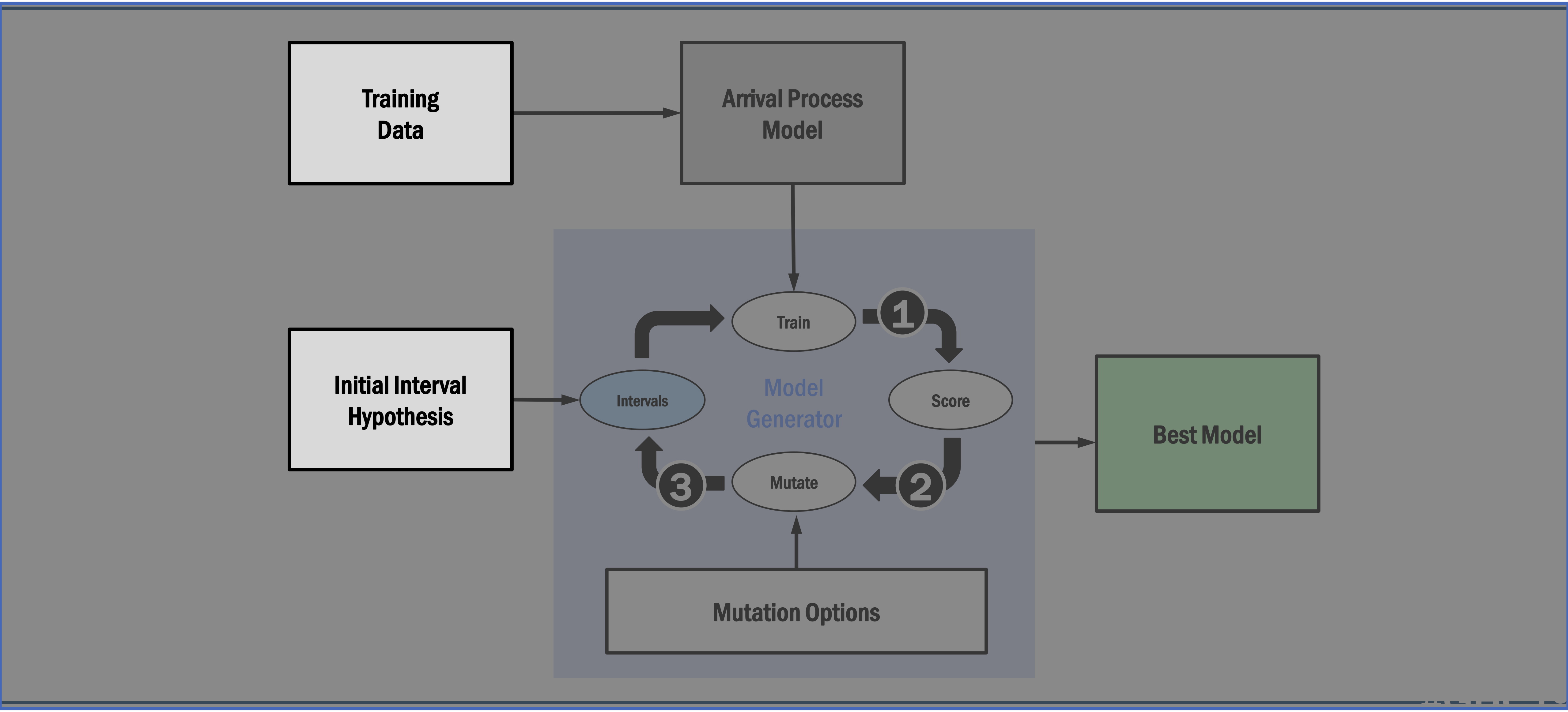
- **Optimization algorithm**
  - **Minimize AICc**
- **Randomly generate a population**
  - **N intervals**
- **Choose the best few intervals to create the next generation with**
  - **Reproduction**
  - **Crossover**
  - **Mutation**
- **Converge after specified number of generations**



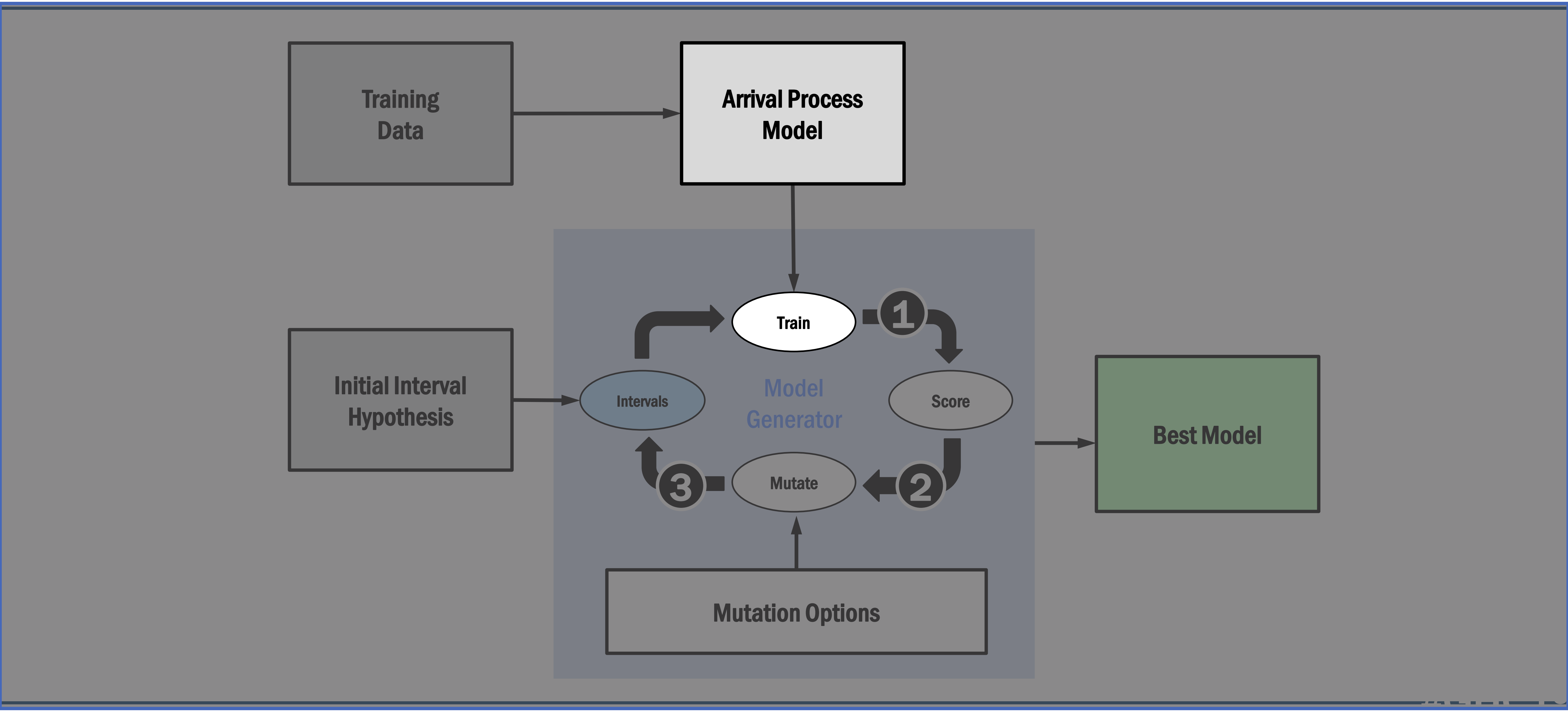
Genetic Algorithm Evolution Flow

Image from AL-Madi, Nagham & Khader, Ahamad Tajudin. (2008). De Jong's sphere model test for a Social-Based Genetic Algorithm (SBGA). IJCSNS. 8.

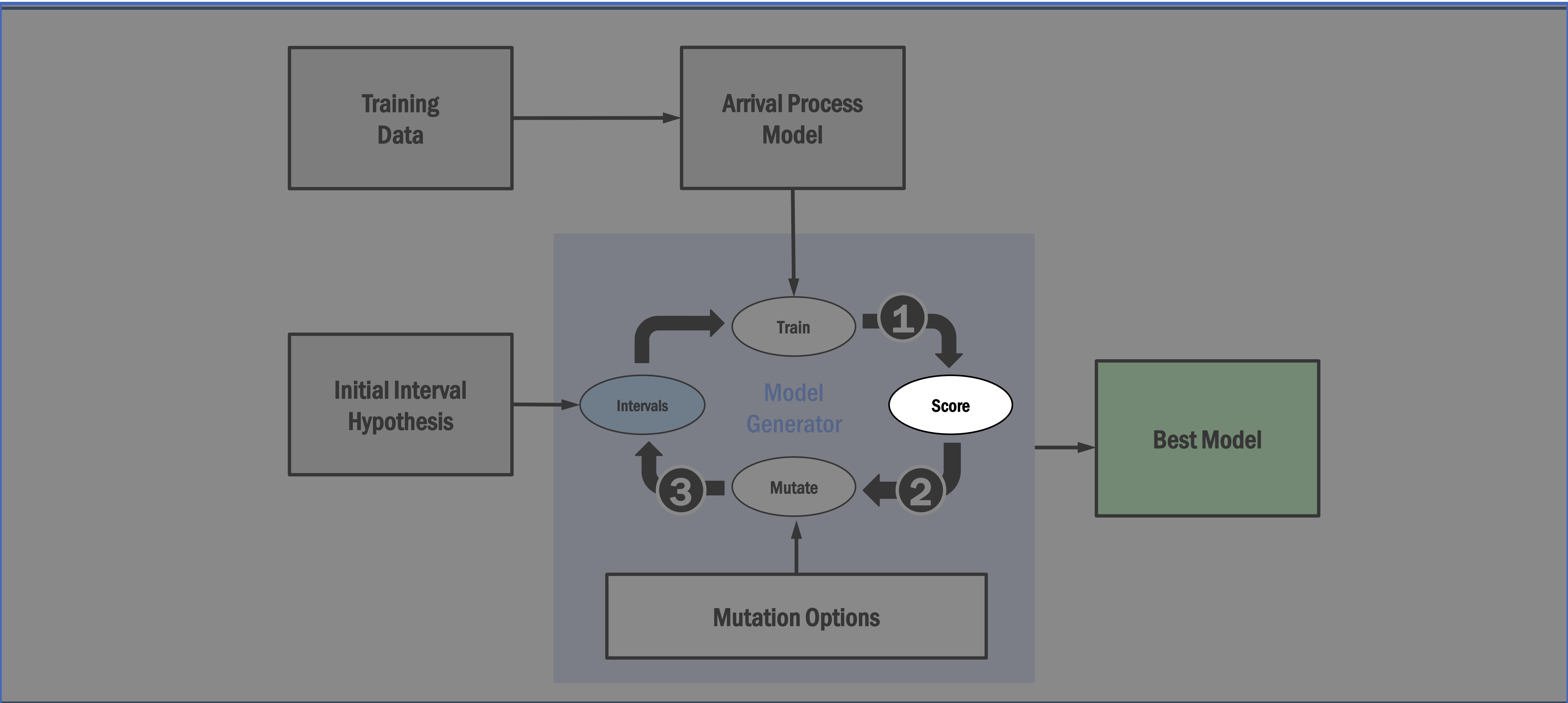
SINAPSE Algorithm



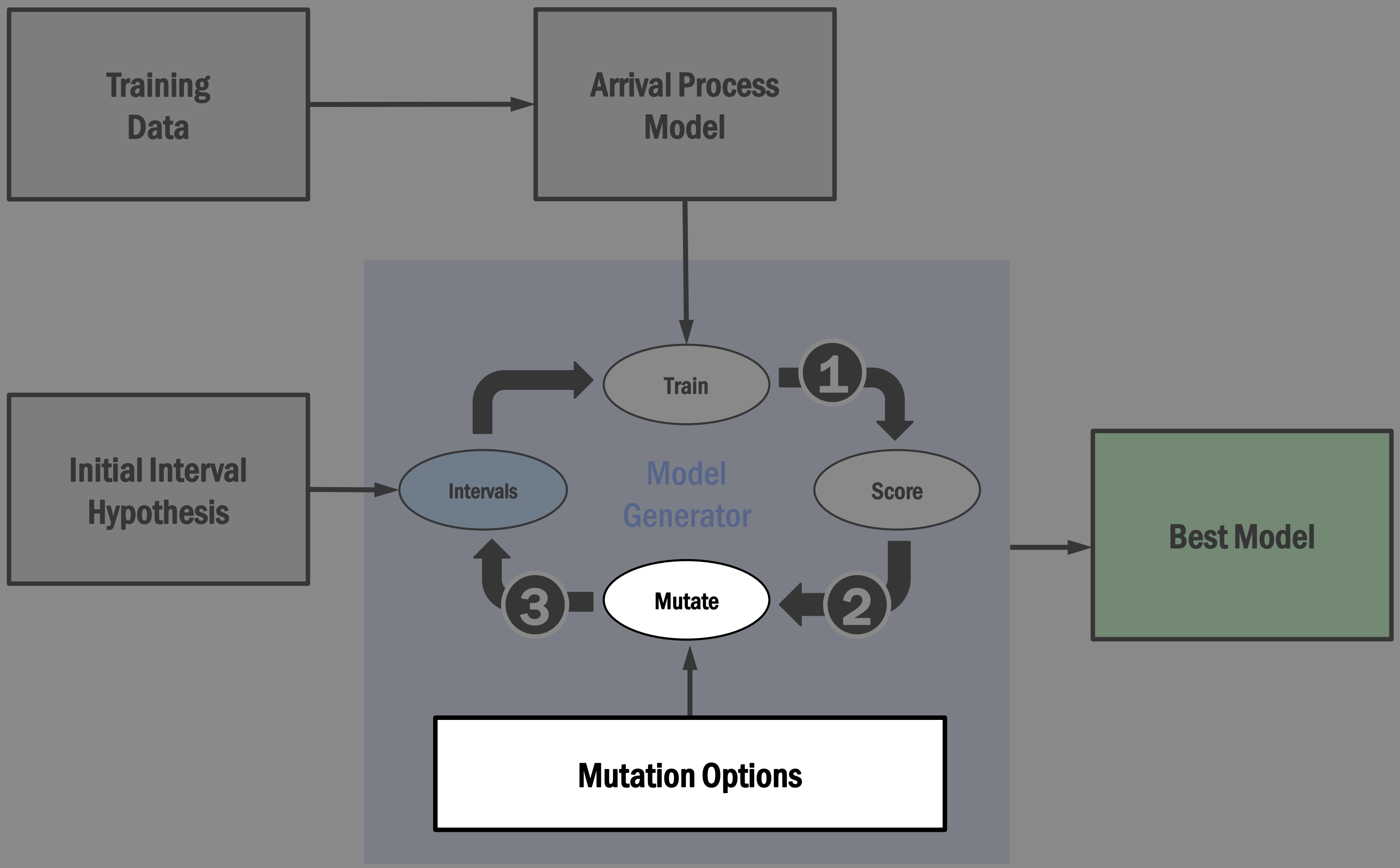
SINAPSE Algorithm



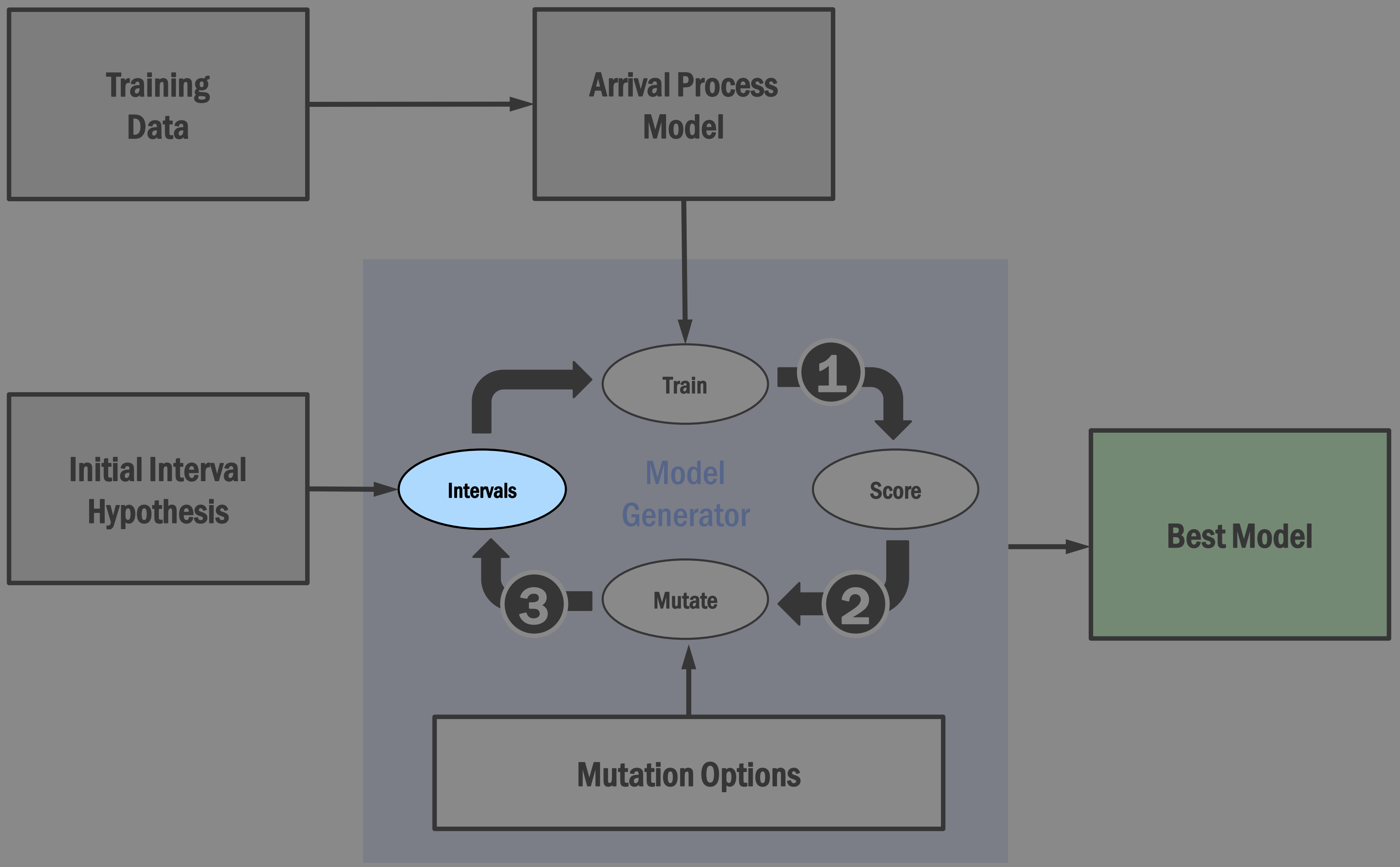
SINAPSE Algorithm



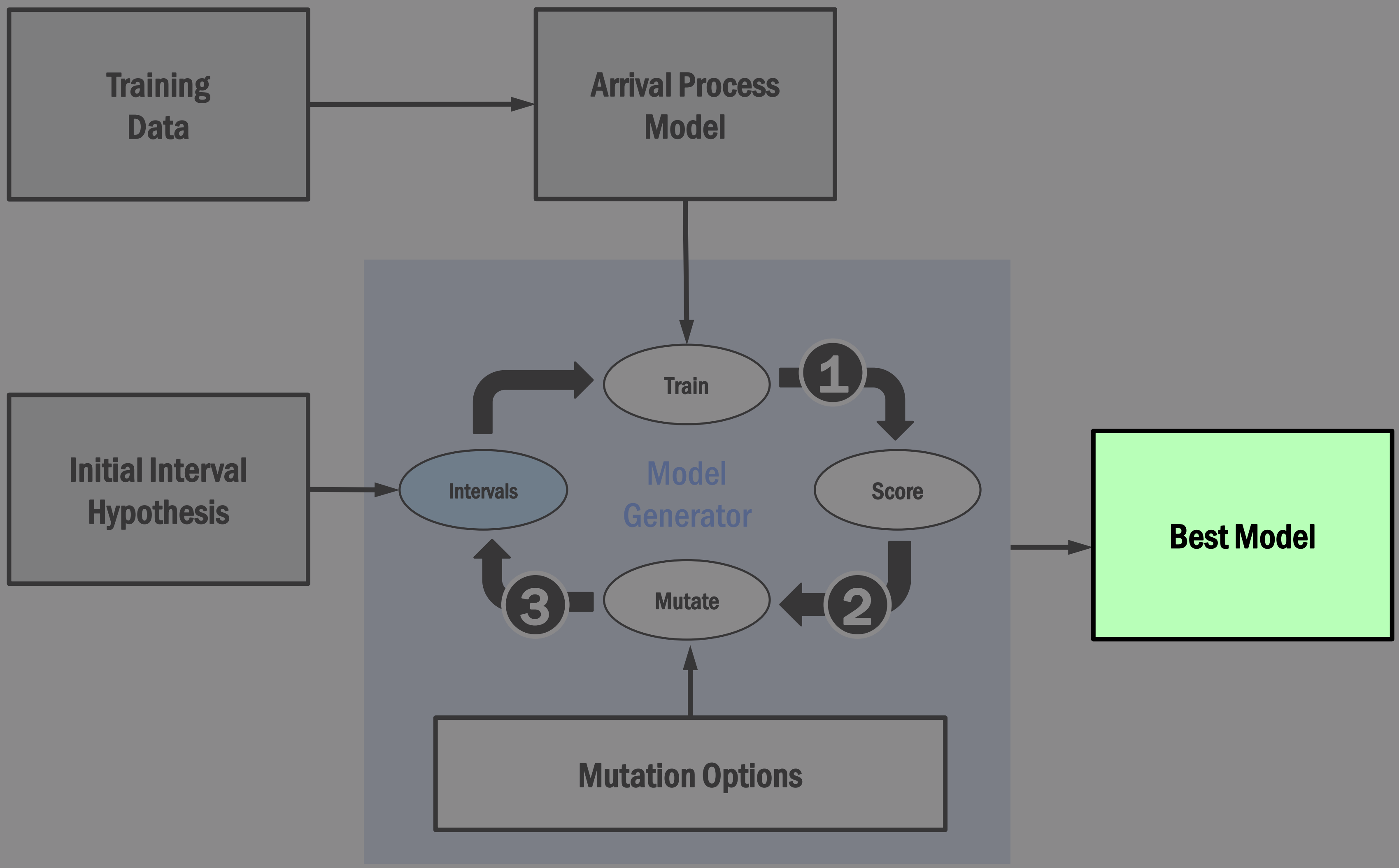
SINAPSE Algorithm



SINAPSE Algorithm



SINAPSE Algorithm



Summary

- **SINAPSE algorithm fits seasonal time series with:**
  - **Minimal parameter setting**
  - **Nonconsecutive intervals**
  - **Simple prediction**
- **When the data has:**
  - **Few samples**
  - **Sharp breakpoints**
  - **Fixed period seasonality**
  - **Many observations per period**
- **It can be used for:**
  - **Modeling seasonality**
  - **Predicting seasonal time series**
  - **Anomaly detection**
- **See our poster and paper for more on applications and results**