PSMC, IICR and SNIF

Lounes Chikhi et Camille Steux

Centre de Recherche sur la Biodiversité et l'Environnement (CRBE) **CNRS, Toulouse, France**
Centre for ecology, evolution and encironmental changes (ce3c) Univ. Lisboa, Portugal
International Research Laboratory (IRP) BEEG-B

Montpellier
DevOCGen, 4 Décembre 2025

Reconstructing the demographic history of species with genetic/genomic data

Other data:
- Linguistics
- Archaeology
- Ecology
- Paleo-climatology

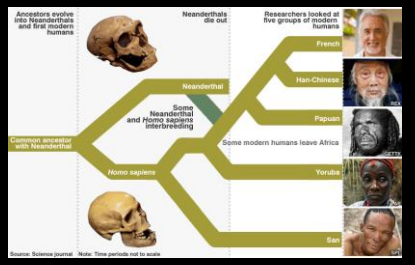
Genetic / genomic data

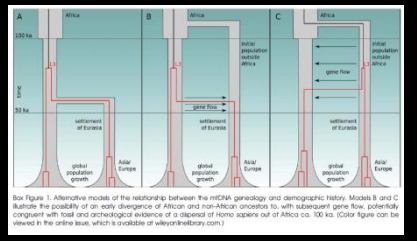
Model ?

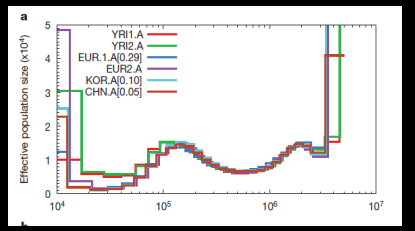
Patterns:
how is diversity distributed ?

Evolutionary history:

- Population size changes
- Population structure
- Selection







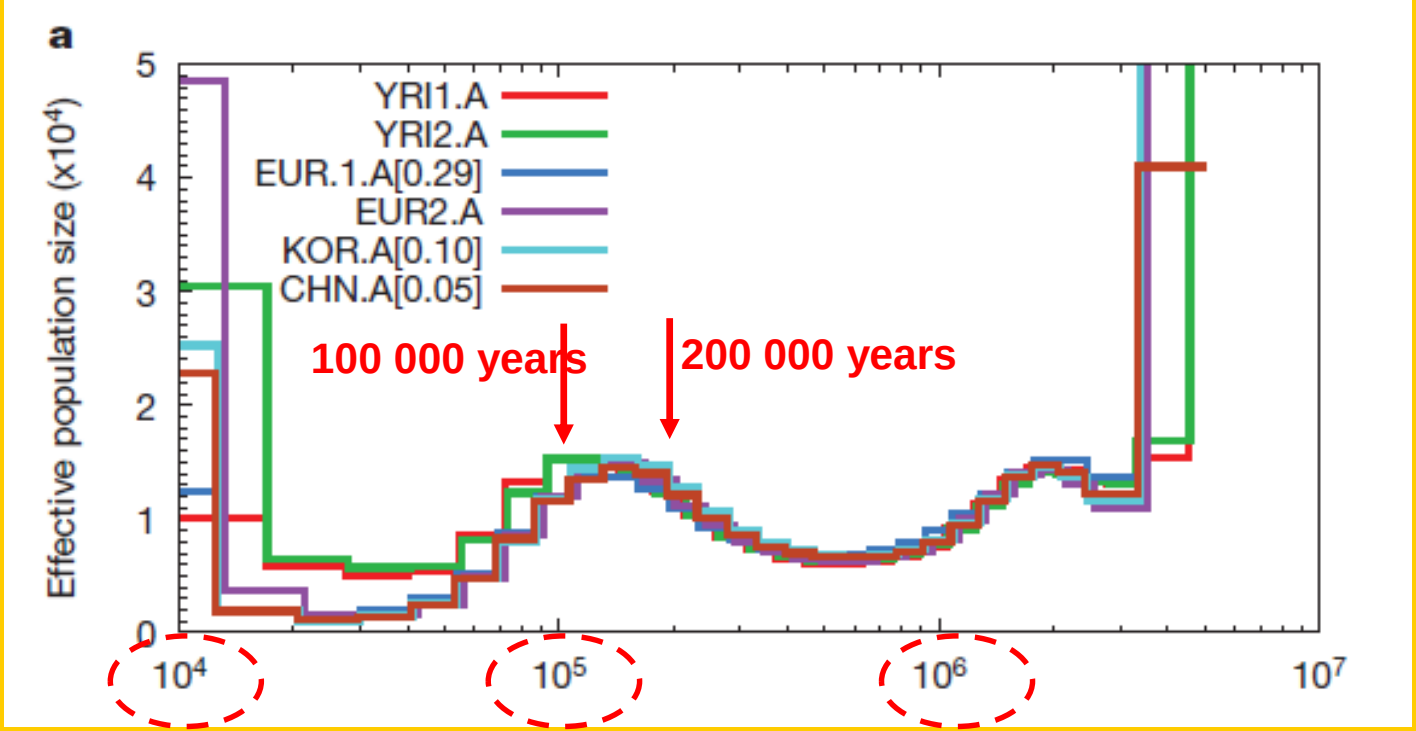
Inference of human population history from individual whole-genome sequences

Heng Li[1,2] & Richard Durbin[1]

NATURE | 2011 |

PSMC

(Pairwise Sequentially Markovian Coalescent)



<-- present **Time (years)** **past**

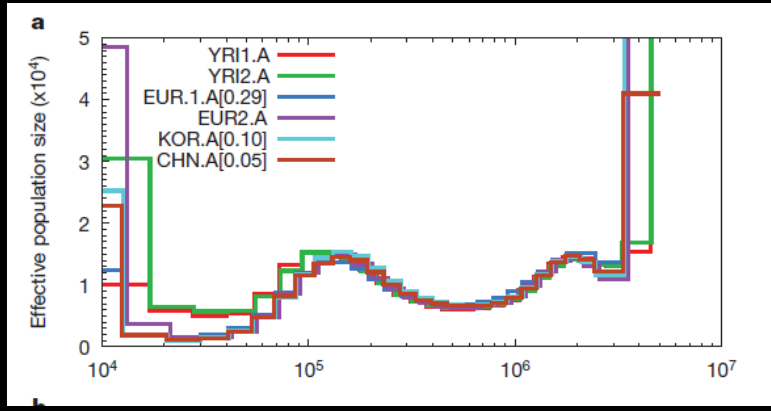
THE PSMC
(Pairwise Sequentially Markovian Coalescent)

Estimates the distribution of coalescence times along the genome of a single diploid individual by looking at the pattern of heterozygous sites.

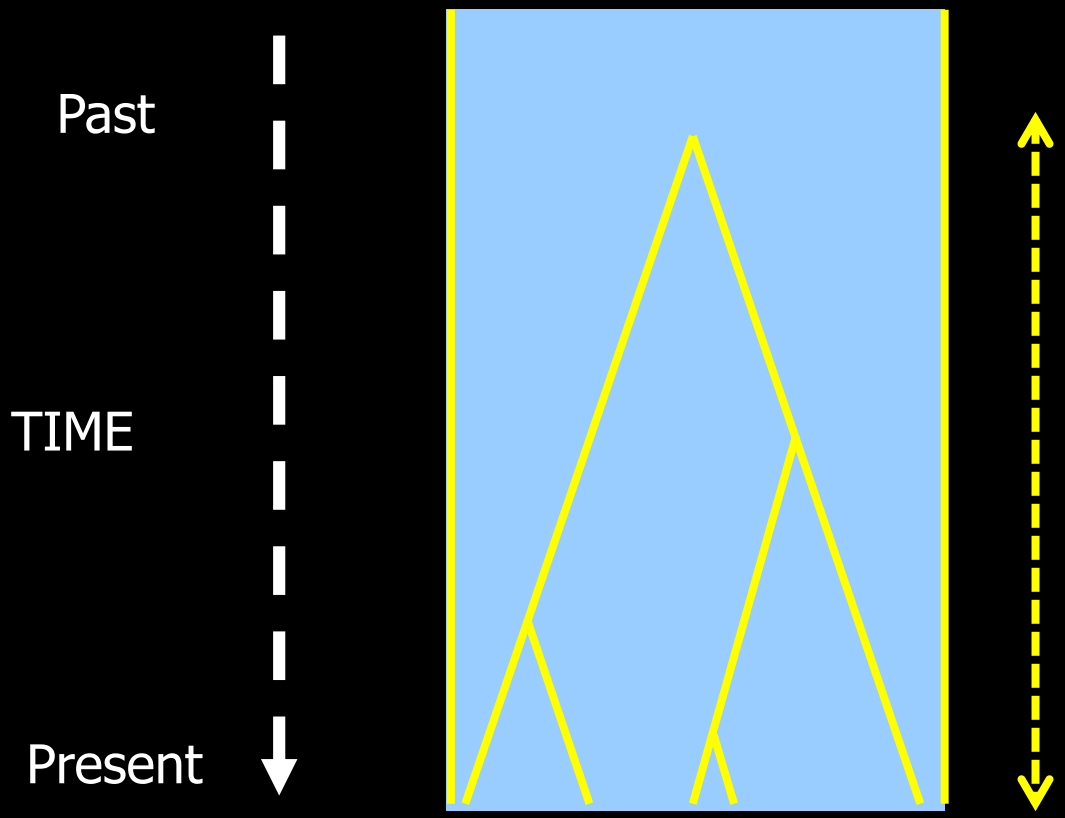
Finds the history of population size changes that best explains this pattern.

Requires
- Mutation rate
- Recombination rate
- Generation time

Assumes panmixia (no population structure / total isolation) in the interpretation step



Coalescence rates in panmictic populations:



Standard Coalescent tree:

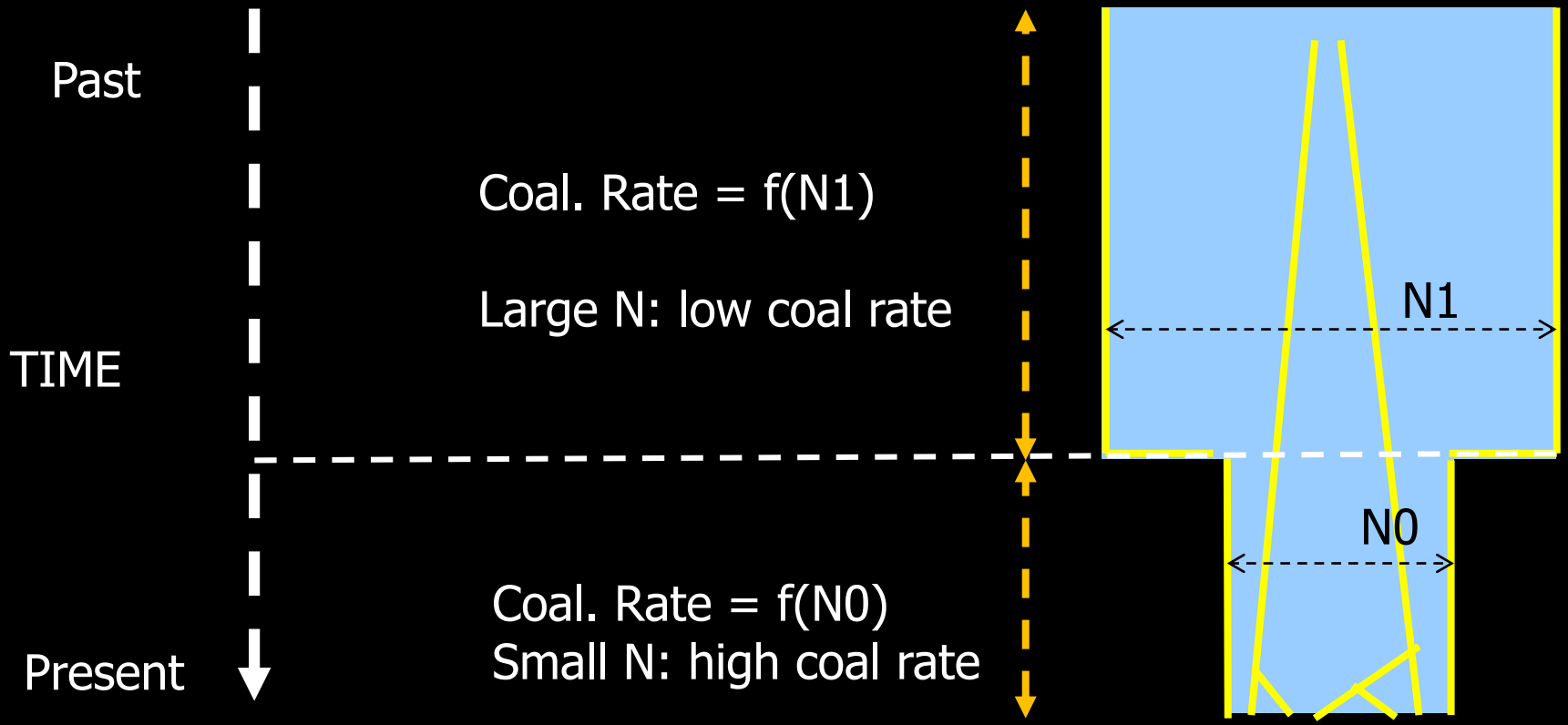
Height = 2N gen. (N haploids)

Coalescence rate k(k-1)/(4N)

1/(2N) for k=2

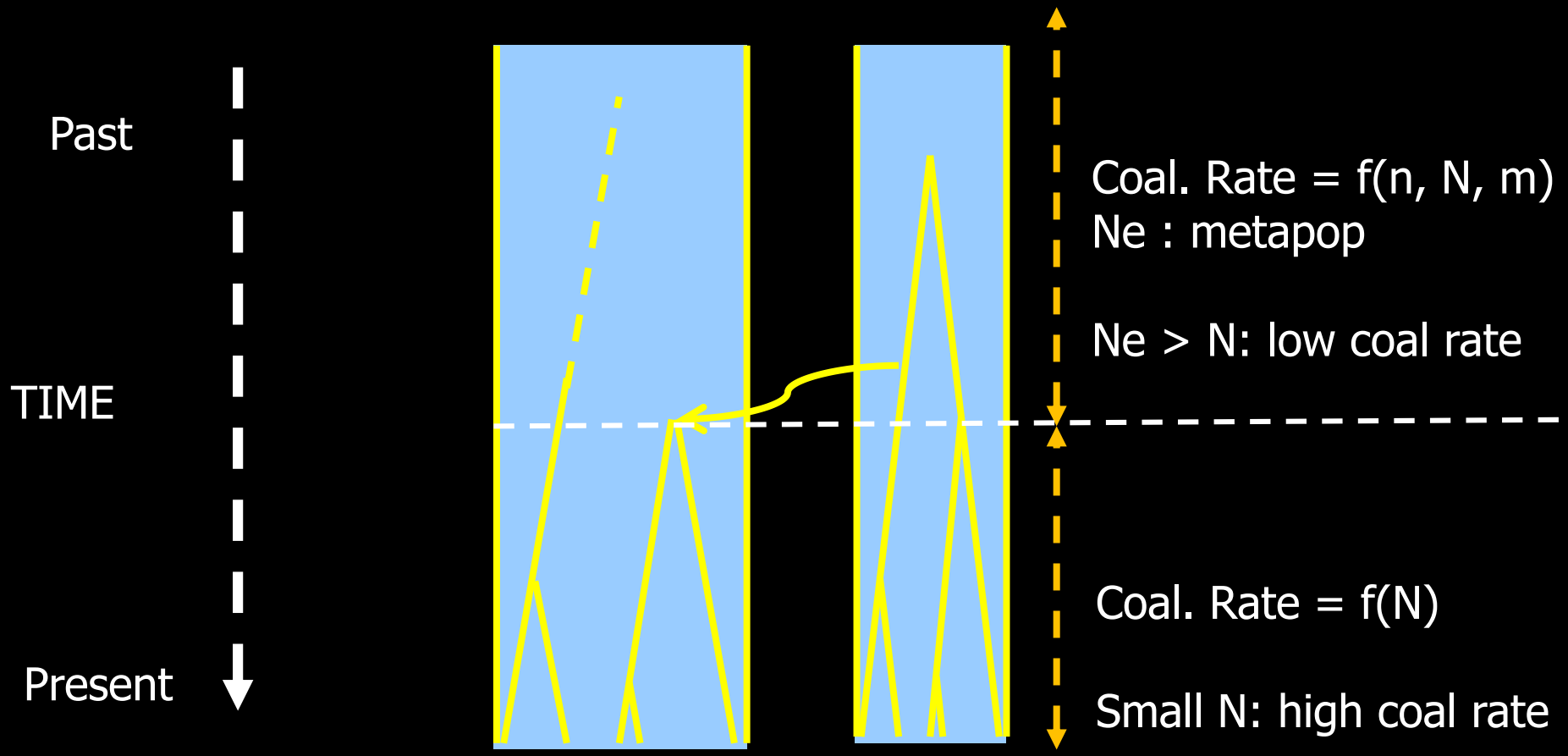
For a diploid k=2

Coalescence rates in panmictic populations: with population size changes

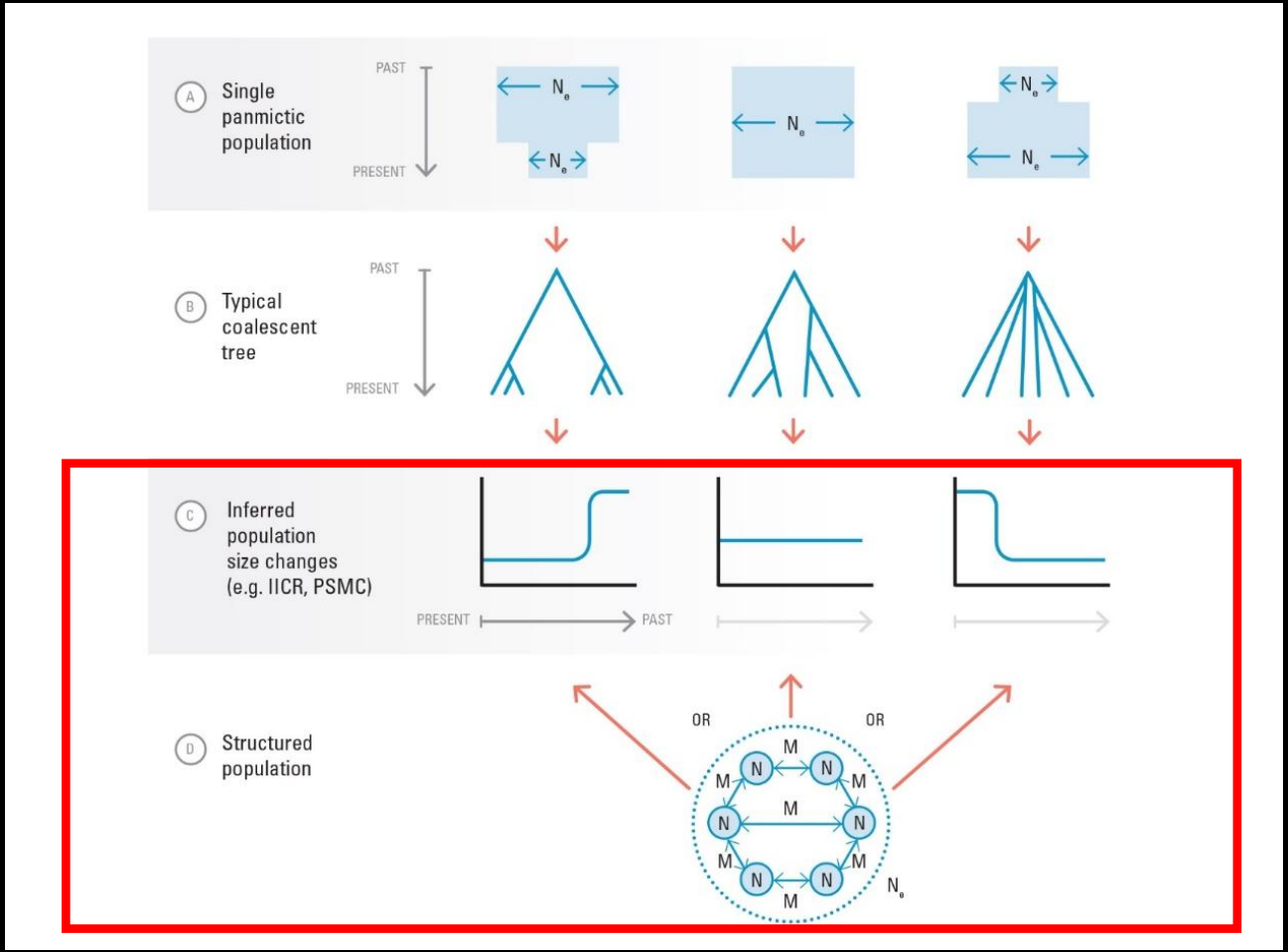


If we can estimate variation in coalescence rate
we can thus estimate variation in N
Example : (Bayesian) Skyline plots, PSMC, MSMC, stairway plots

Departures from the panmictic model
population structure



Population size change *versus* population structure



Did Our Species Evolve in Subdivided Populations across Africa, and Why Does It Matter?

Scerri et al. 2018

Trends in Ecology & Evolution

The IICR
(inverse instantaneous coalescence rate)

The IICR can be seen as a (fictitious) history of population size change that will perfectly fit the distribution of coalescence times (T_2) that will be inferred from a single genome, even if the individual comes from a population that is not panmictic.

It is the inferred history IF the population is panmictic

Otherwise (population structure) this (fictitious) history is the IICR of that sample (IICR_2)

The IICR is the inverse of a coalescence rate, and thus is equivalent to a Ne in a panmictic population

IICR = f(Phi, [s], t) Phi = model parameters

ORIGINAL ARTICLE

On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference?

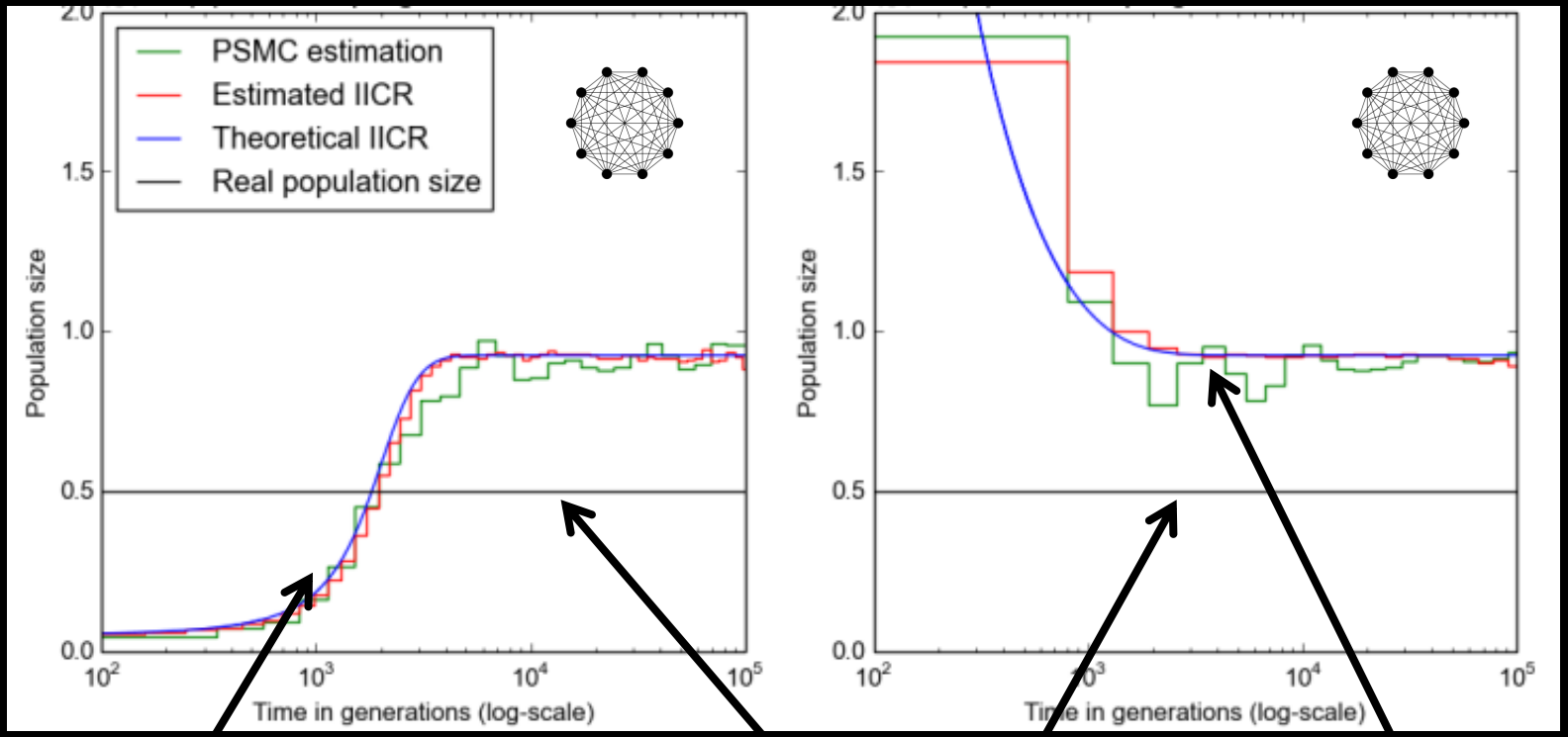
O Mazet[1,7], W Rodríguez[1,7], S Grusea[1], S Boitard[2,3] and L Chikhi[4,5,6]

The IICR_2 for the n-island model

sampling

same island

two different islands



signal of spurious pop size decrease

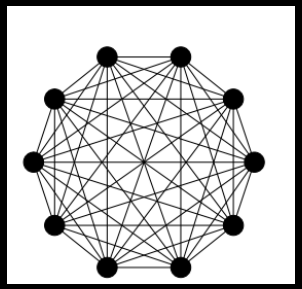
constant size of the n-island

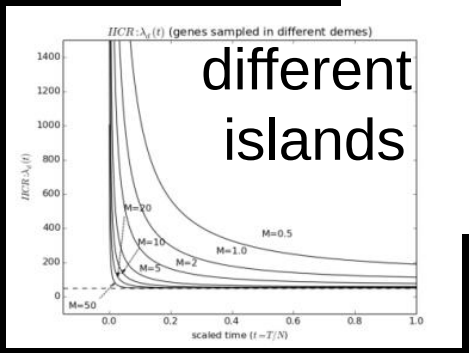
signal of spurious pop size increase

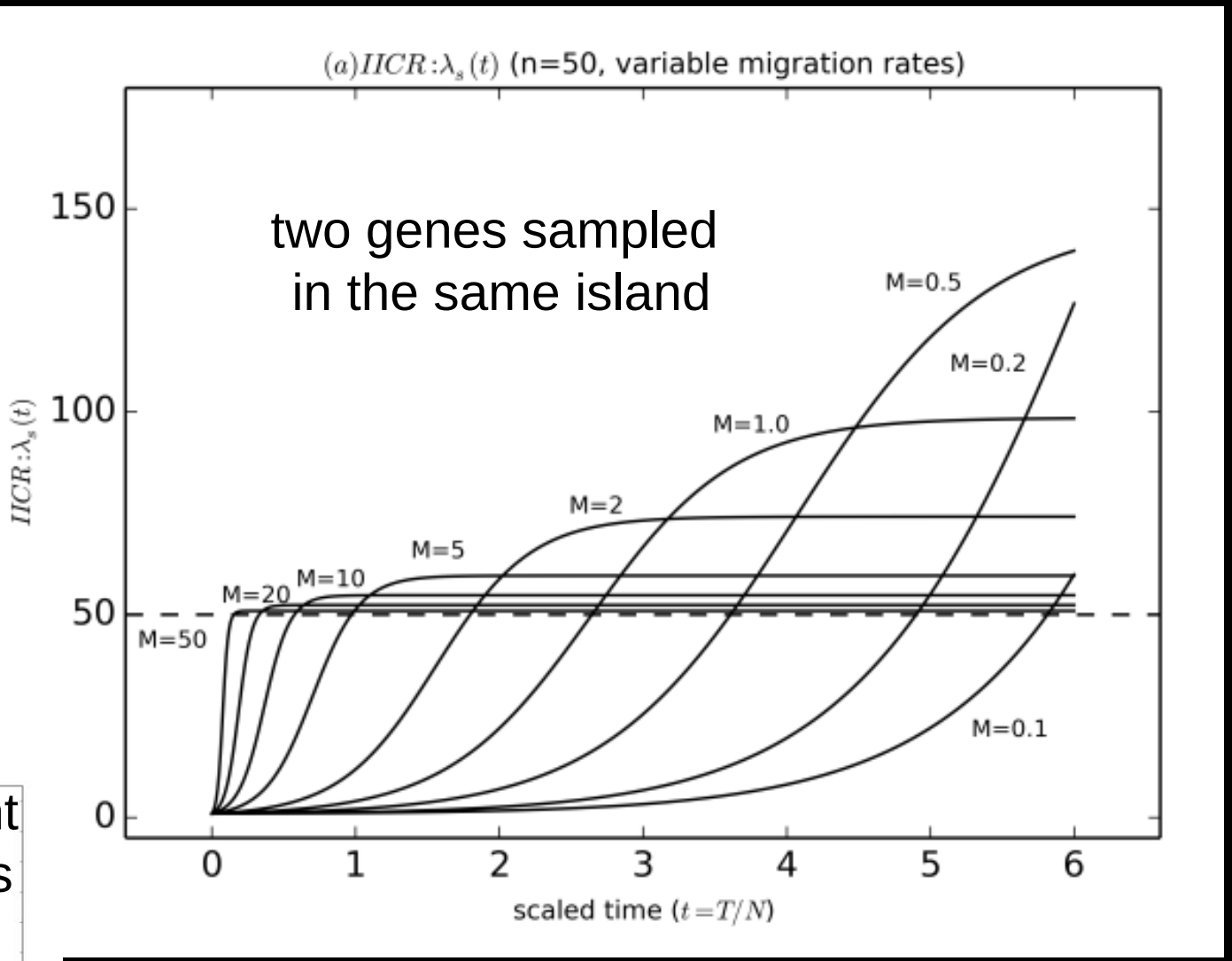
Comparison with the PSMC:
Pop structure
NO population size change

The IICR_2 for the n-island model

n-island
n=50





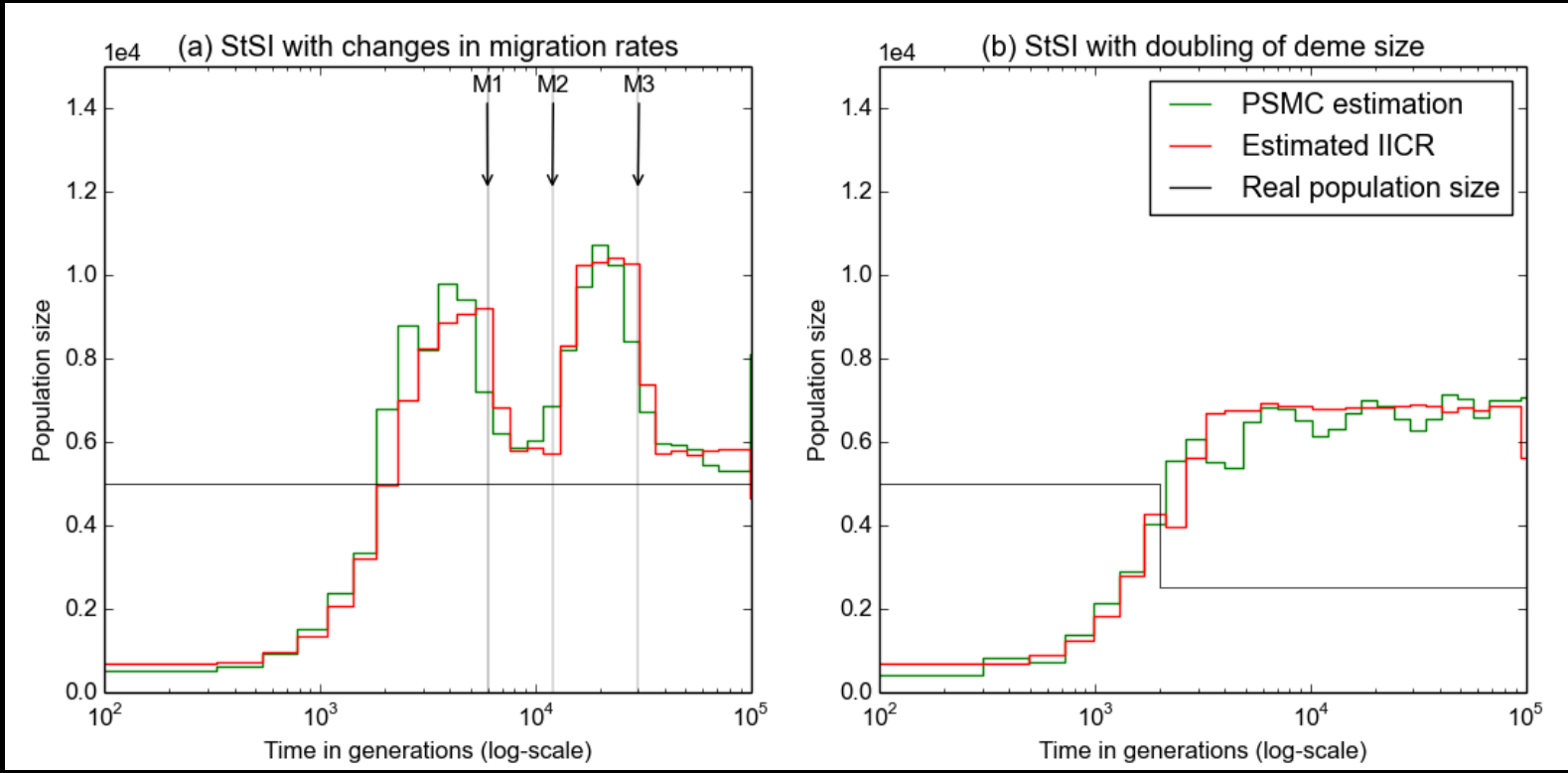


n = 50 islands
M: number of migrants

The IICR_2 for the n-island model

Pop structure
Change in gene flow
No Pop size change

Pop structure
No change in gene flow
Increase in pop size

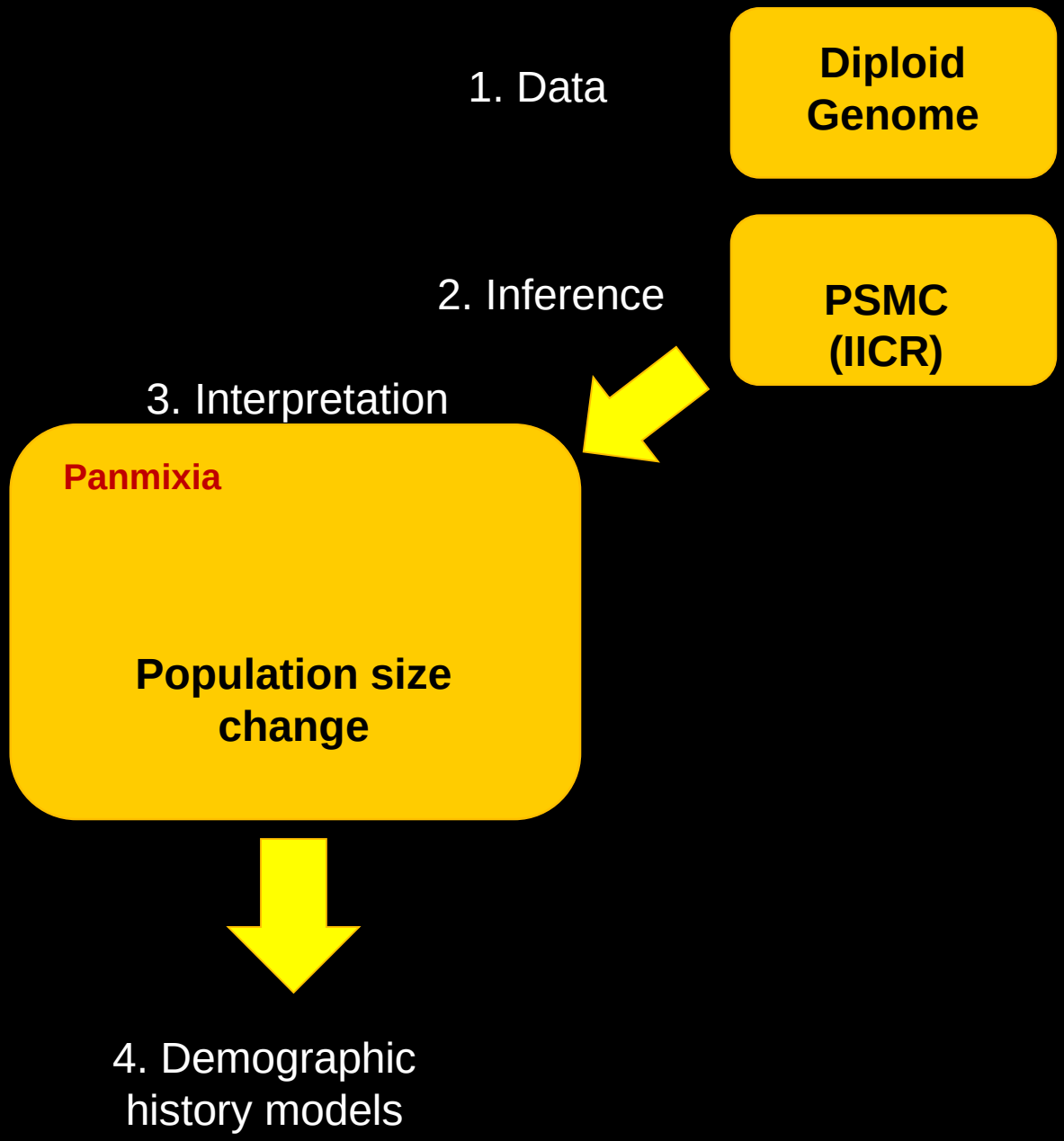


M1=1 – M2=5 – M3=0.8 – M=5 migrants per gen.
T=6,000 – T2= 12,000 T3=30,000 generations

ORIGINAL ARTICLE

On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference?

O Mazet[1,7], W Rodríguez[1,7], S Grusea[1], S Boitard[2,3] and L Chikhi[4,5,6]


1. Data
Diploid Genome
2. Inference
PSMC (IICR)
3. Interpretation
Panmixia
Population size change
4. Demographic history models

1. Data

Diploid Genome

2. Inference

PSMC (IICR)

3. Interpretation

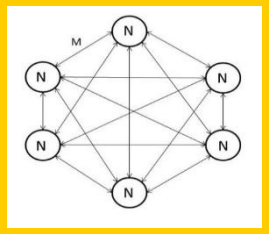
Structure with or without population size change

IICR

SNIF

Manual estimation

n-island model



Number of islands *n*
Migration *M*
Events: change of M at time t_i

Stepping-stone
Split
Continent-island
Pop size change
etc…

4. Demographic history models

Inferring number of populations and changes in connectivity under the n-island model

Armando Arredondo[1,2] · Beatriz Mourato[2,3] · Khoa Nguyen[1] · Simon Boitard[4] · Willy Rodríguez[2,5] · Camille Noûs[6] · Olivier Mazet[1,2] · Lounès Chikhi[3,7]

1. Data

Diploid Genome

2. Inference

PSMC (IICR)

3. Interpretation

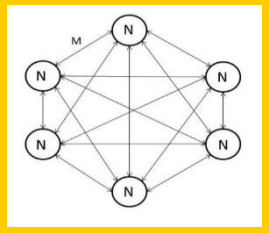
Structure with or without population size change

IICR

SNIF

Manual estimation

n-island model



Number of islands *n*
Migration *M*
Events: change of M at time t_i

Stepping-stone
Split
Continent-island
Pop size change
etc…

4. Demographic history models

1. Data

Diploid Genome

2. Inference

PSMC (IICR)

3. Interpretation

Panmixia

Population size change

4. Demographic history models

3. Interpretation

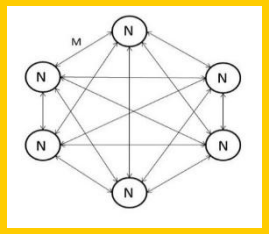
Structure with or without population size change

IICR

SNIF

Manual estimation

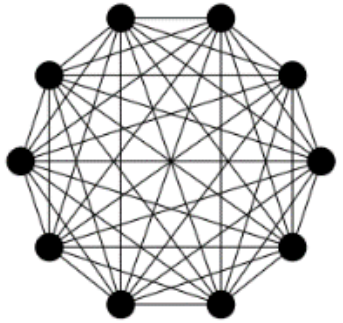
n-island model



Number of islands *n*
Migration *M*
Events: change of M at time t_i

Stepping-stone
Split
Continent-island
Pop size change
etc…

SNIF: Structured Non-stationary Inferential Framework

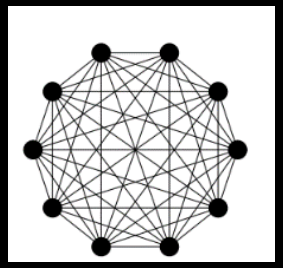


- Inference of demographic parameters
- Differential Evolution as the optimization technique
- Distance used between IICR curves:

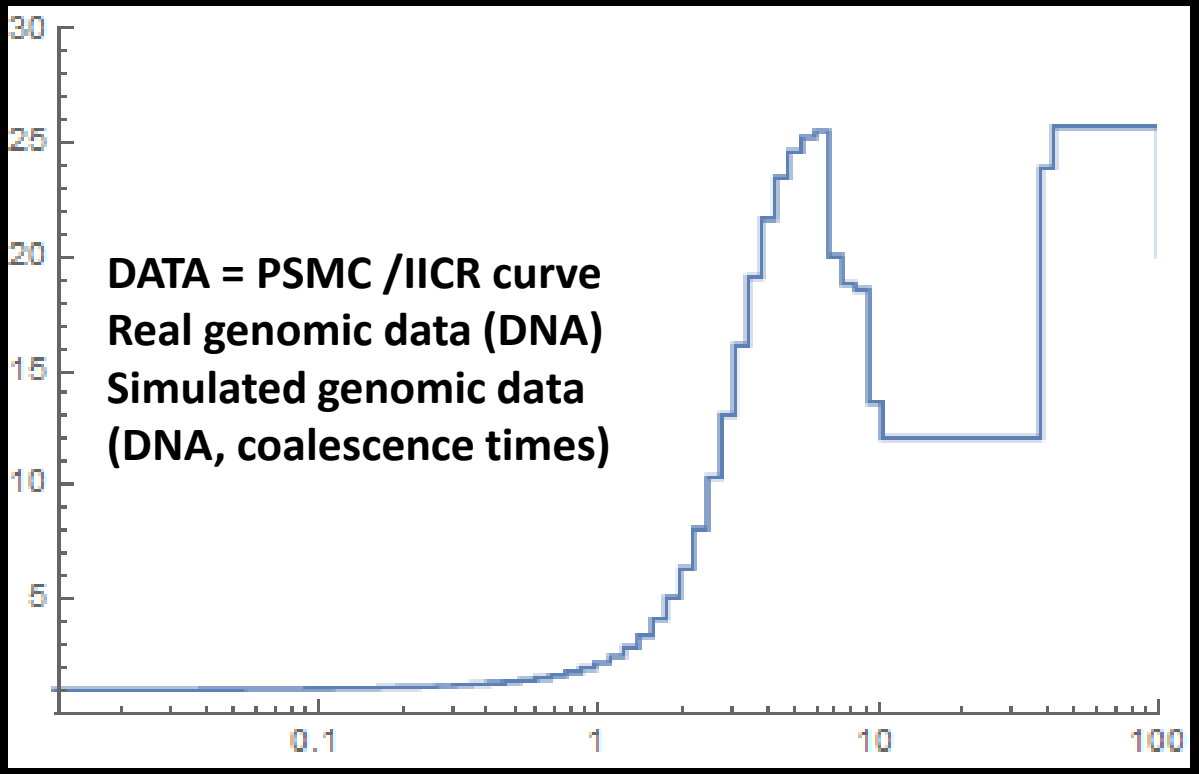
$$\mathrm{d}(\mathrm{IICR}_0, \mathrm{IICR}_1) = \int_0^\infty \left|\mathrm{IICR}_0(t) - \mathrm{IICR}_1(t)\right| \cdot pdf_{\mathrm{IICR}_1}(t)\ \mathrm{d}t$$

- Implementation using Python SciPy library + D Programming Language for performance-sensitive parts

n = nb of islands
M0, M1, etc. Migration
t0, t1, t2, etc. Time of change



IICR



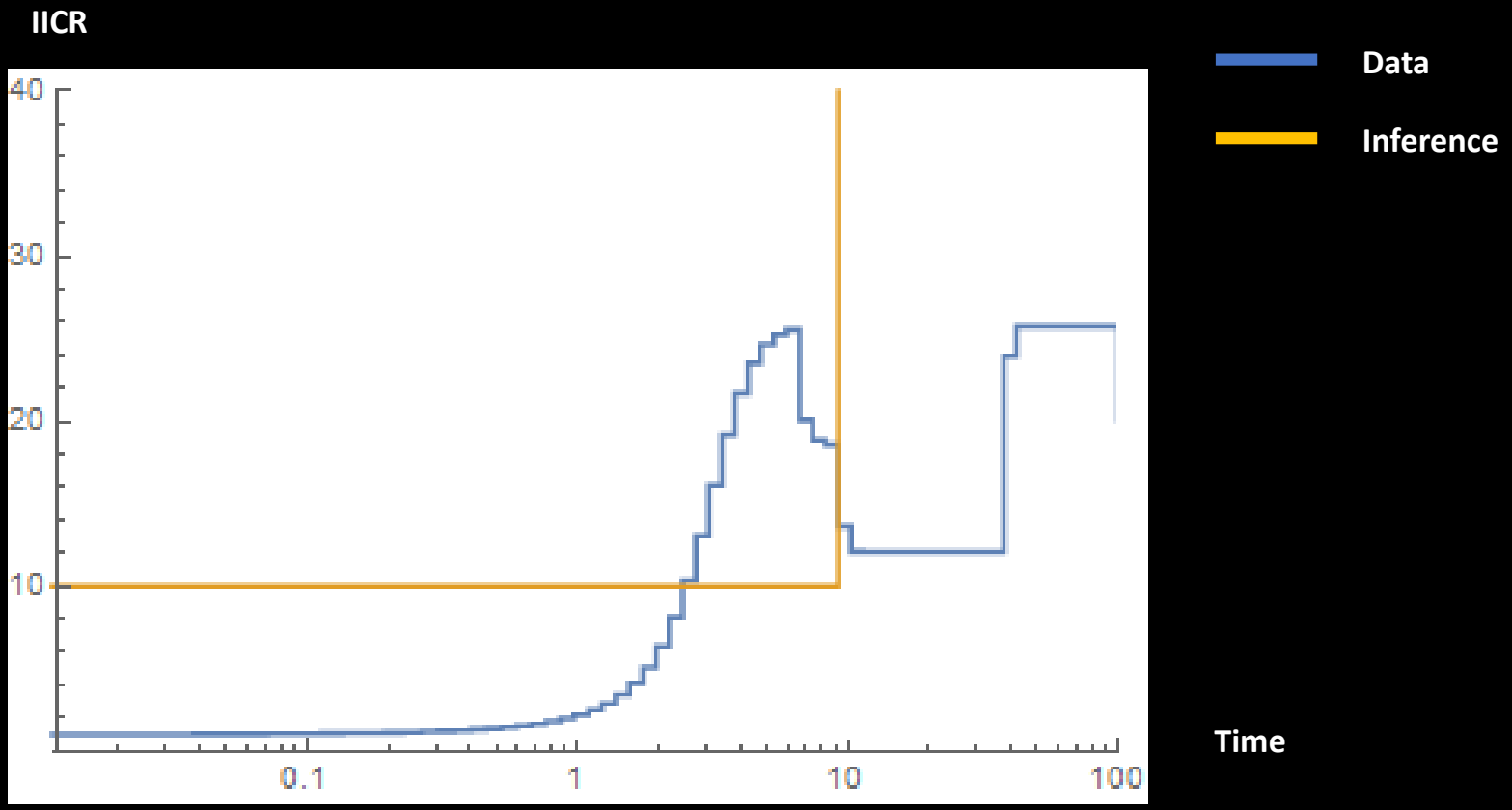
Data

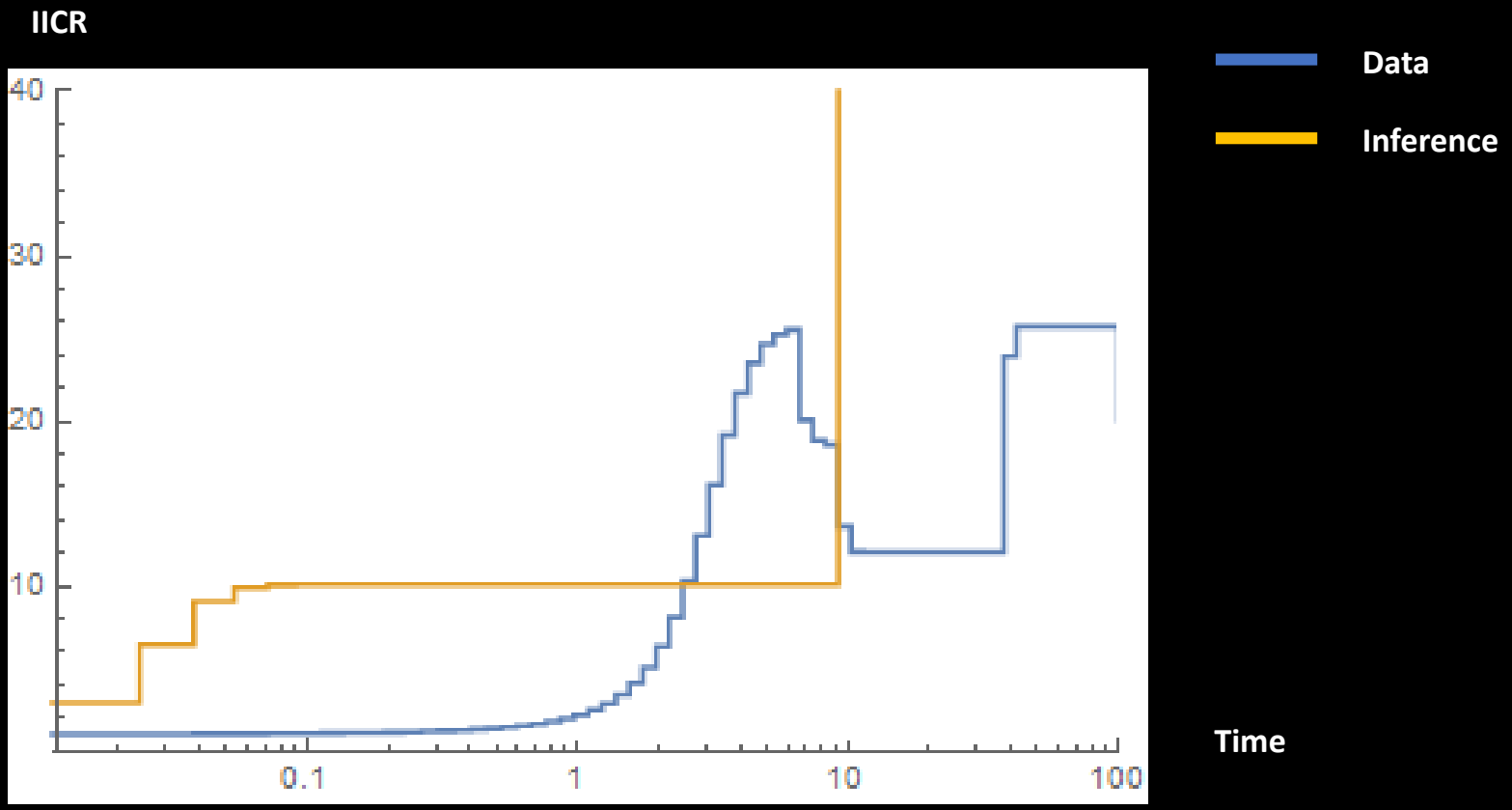
Time

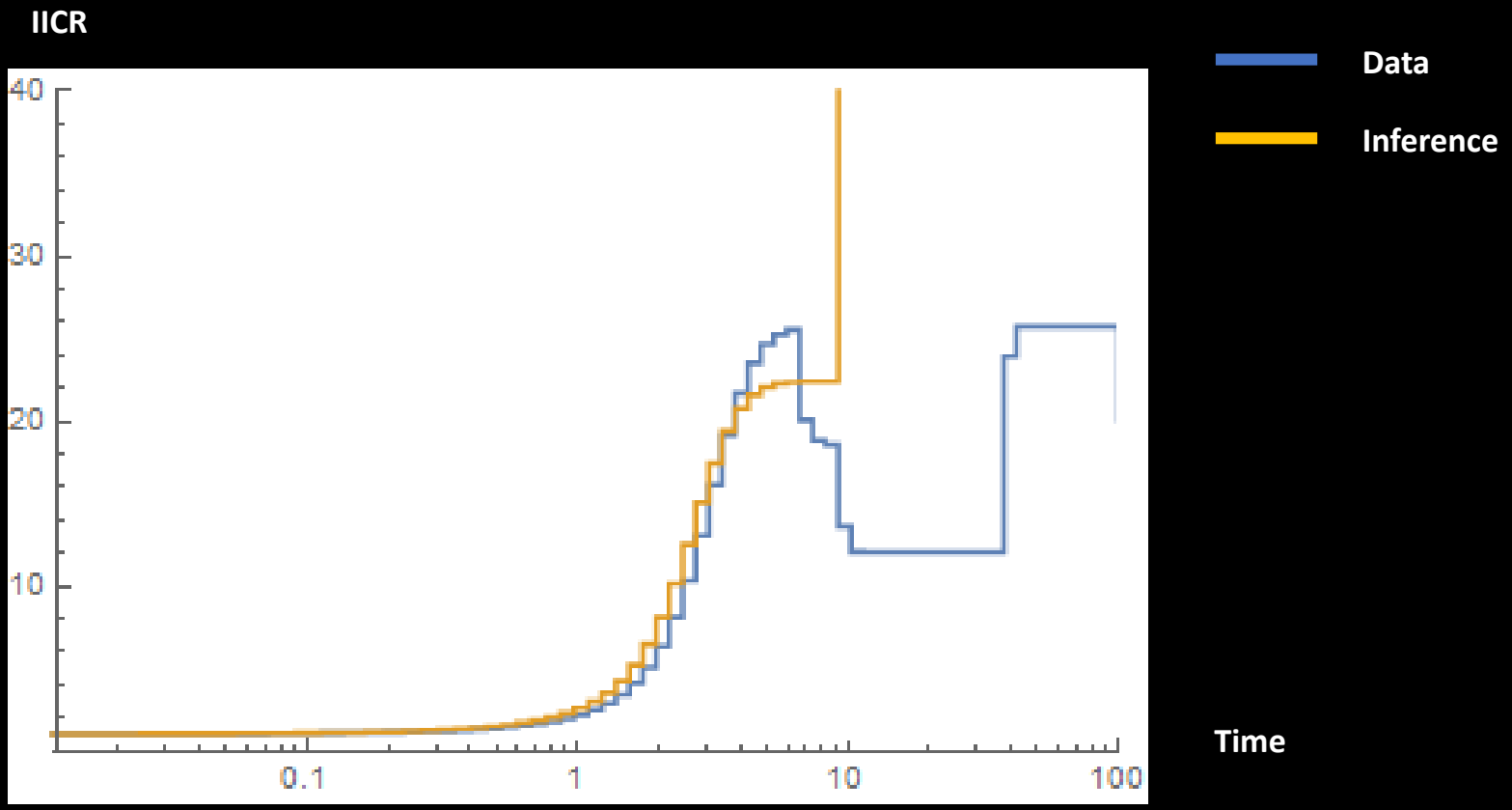
present

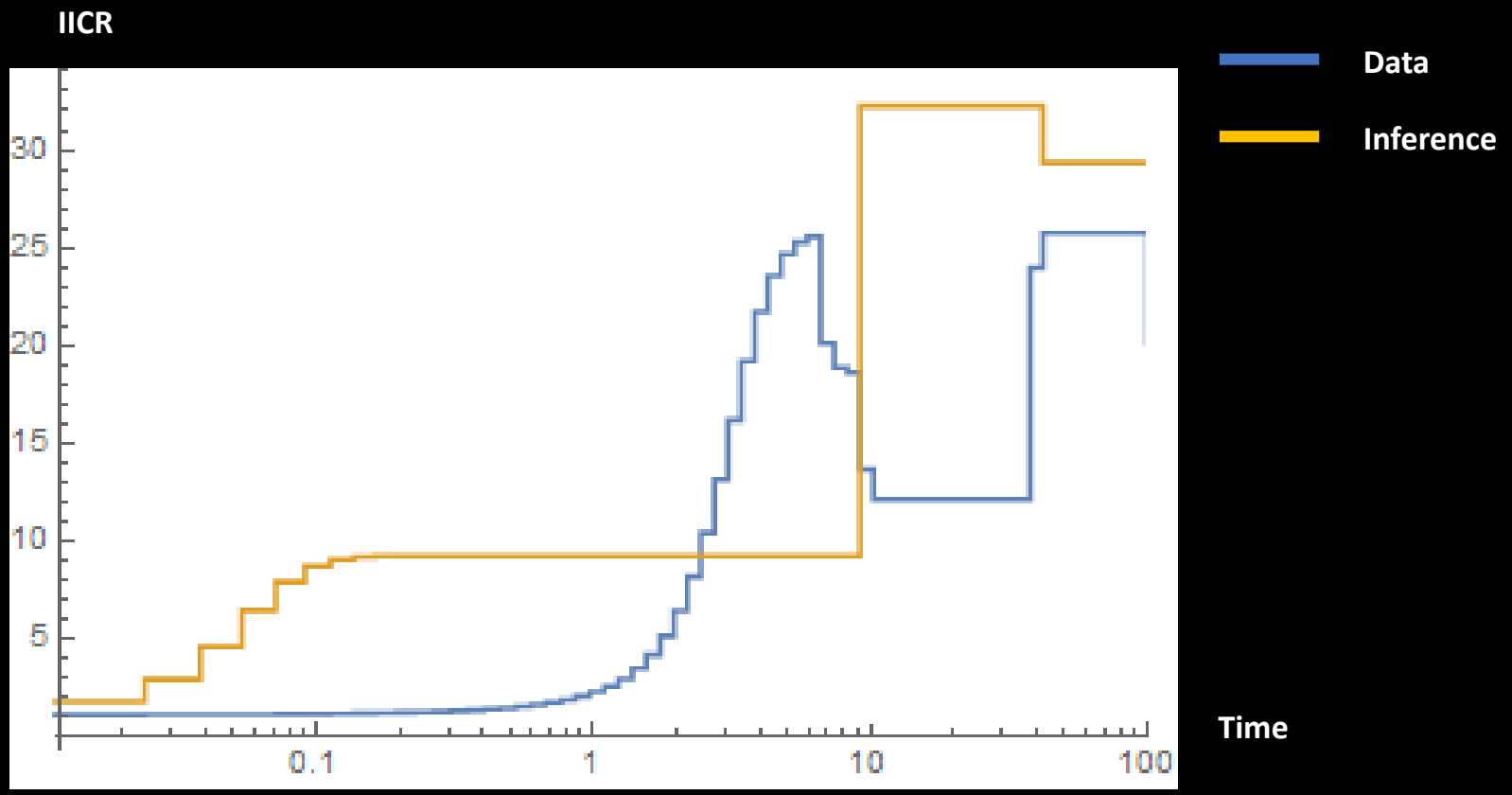
past

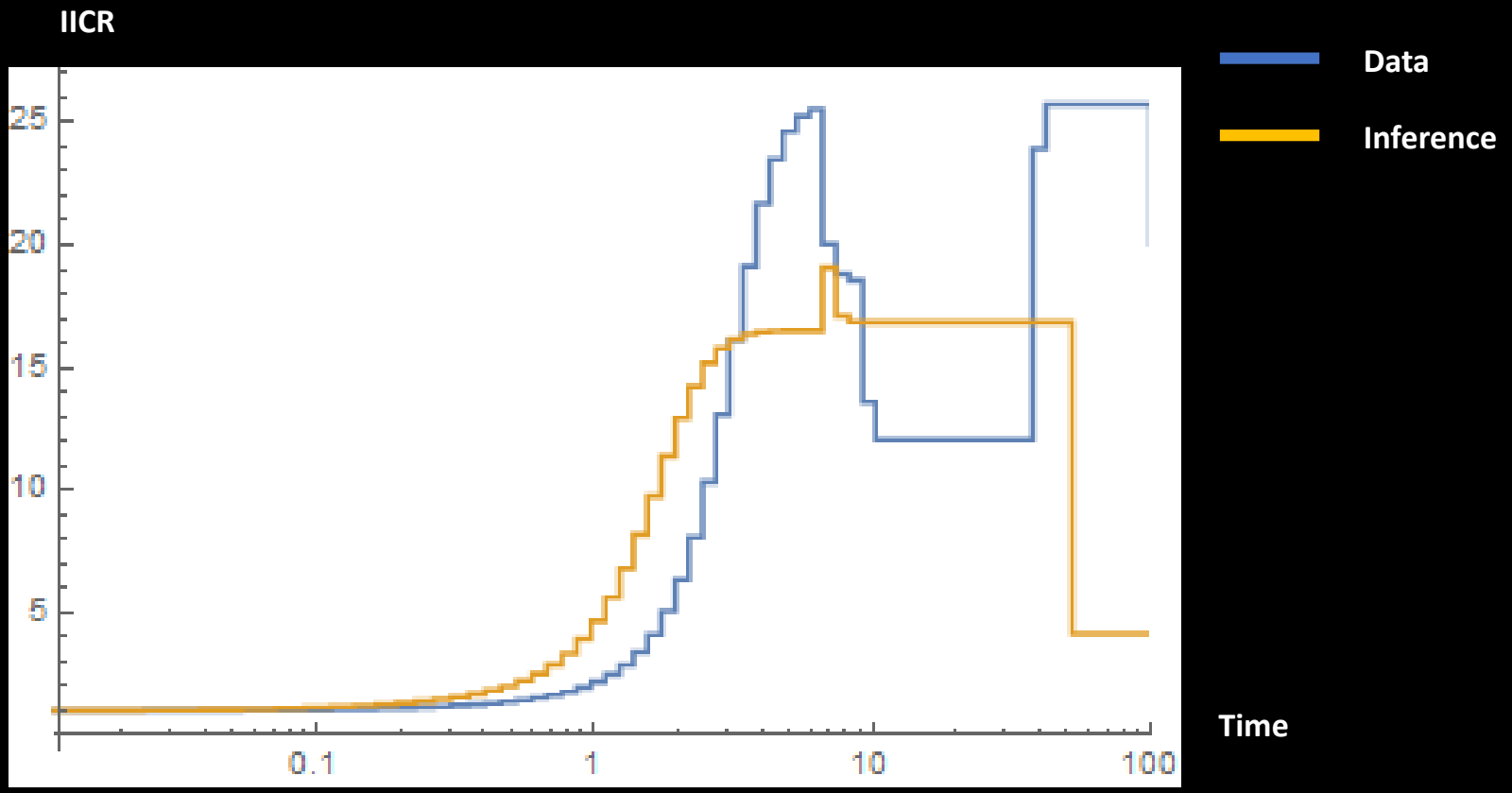
Time

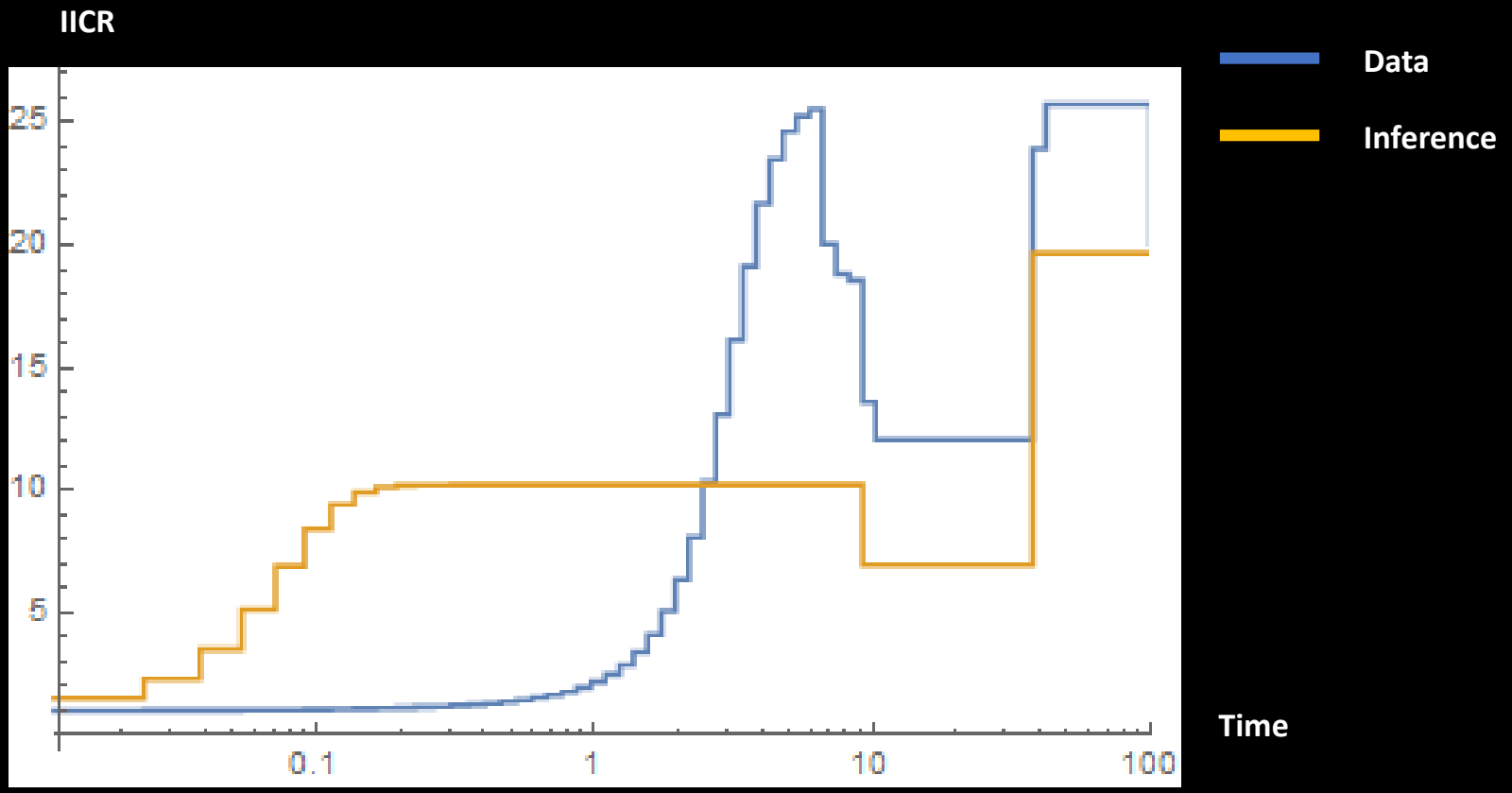

IICR
40
30
20
10
0.1
1
10
100
Data
Inference
Time

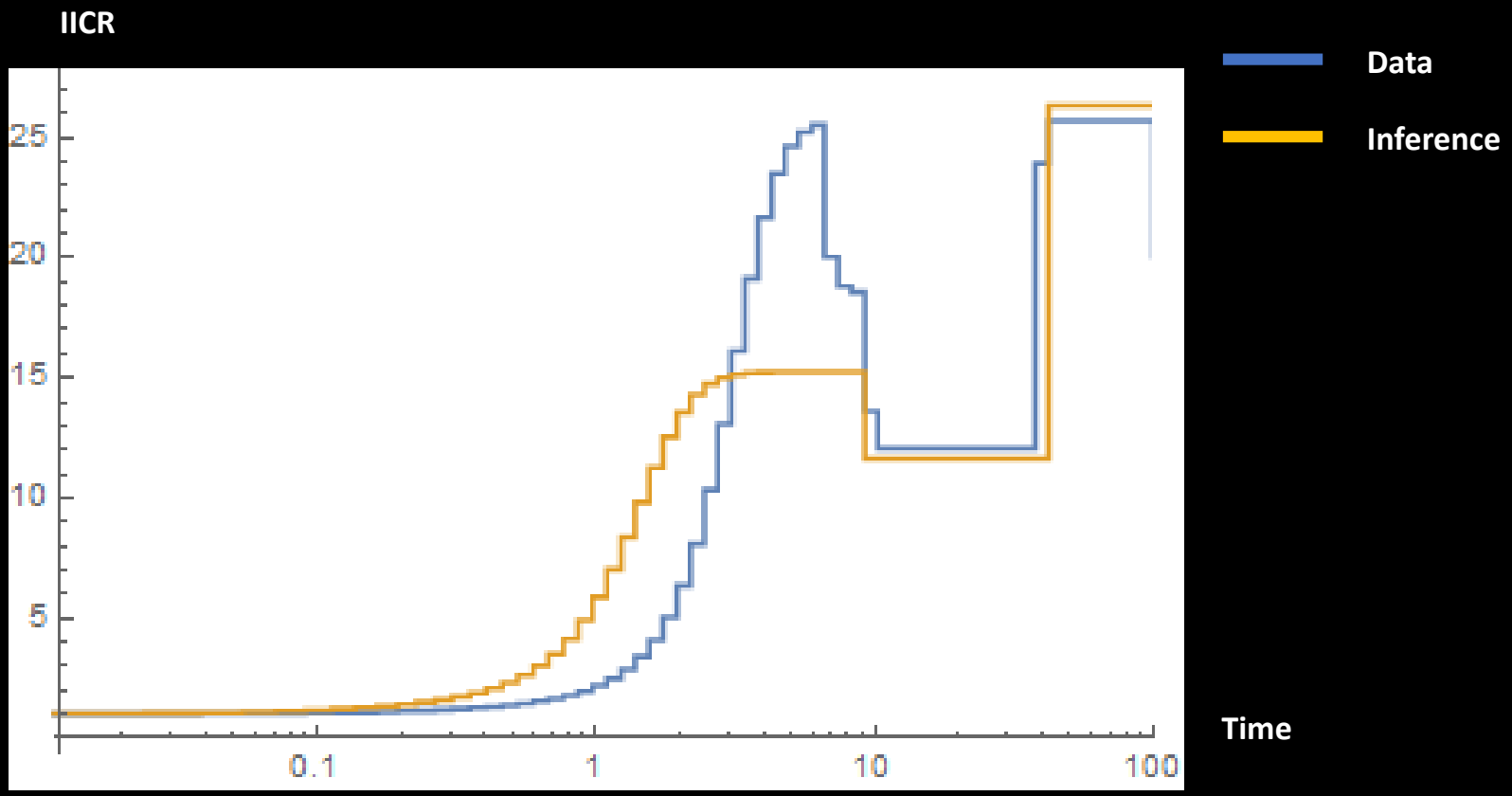

IICR
40
30
20
10
0.1
1
10
100
Data
Inference
Time

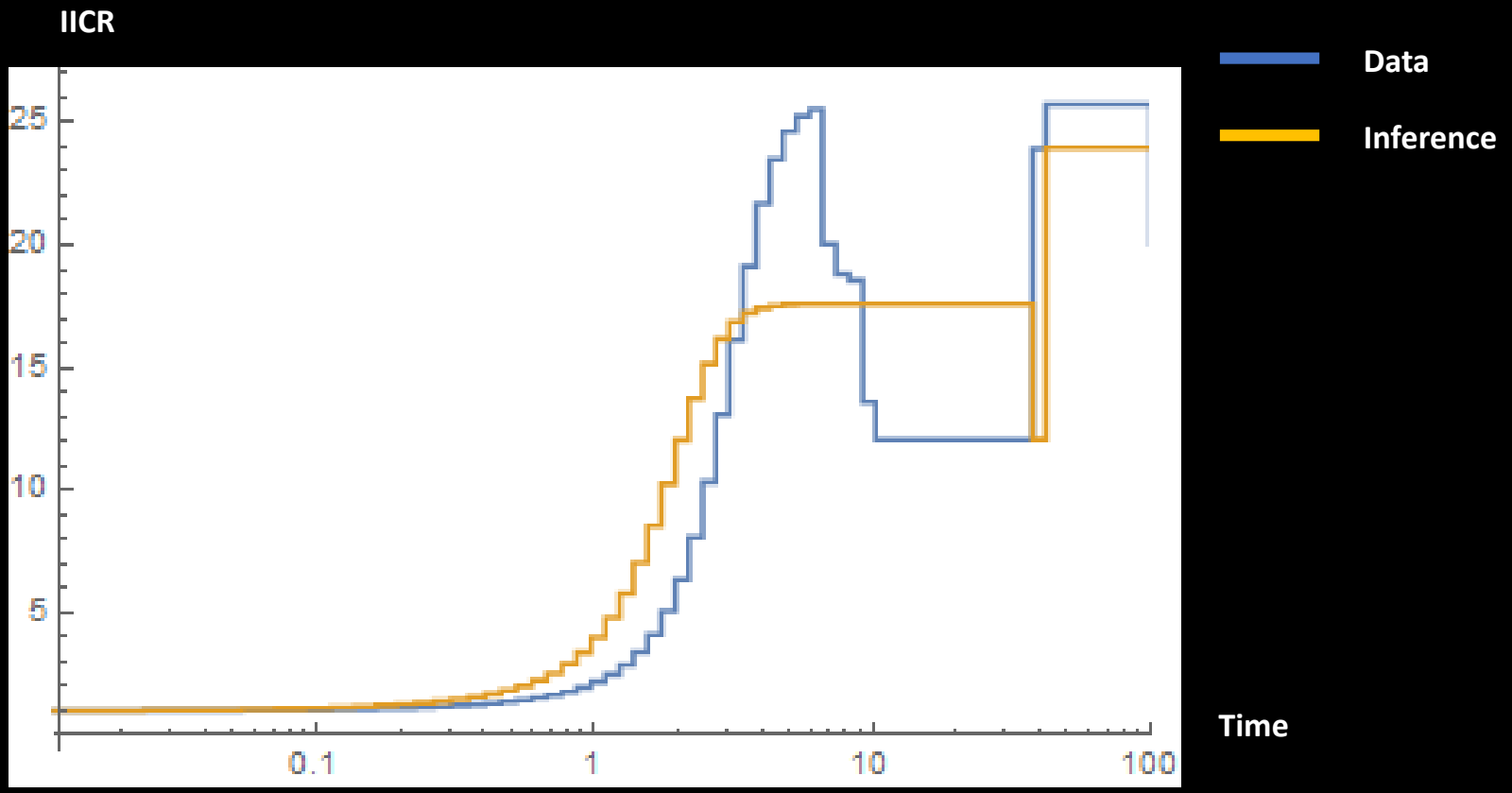

IICR
40
30
20
10
0.1
1
10
100
Data
Inference
Time

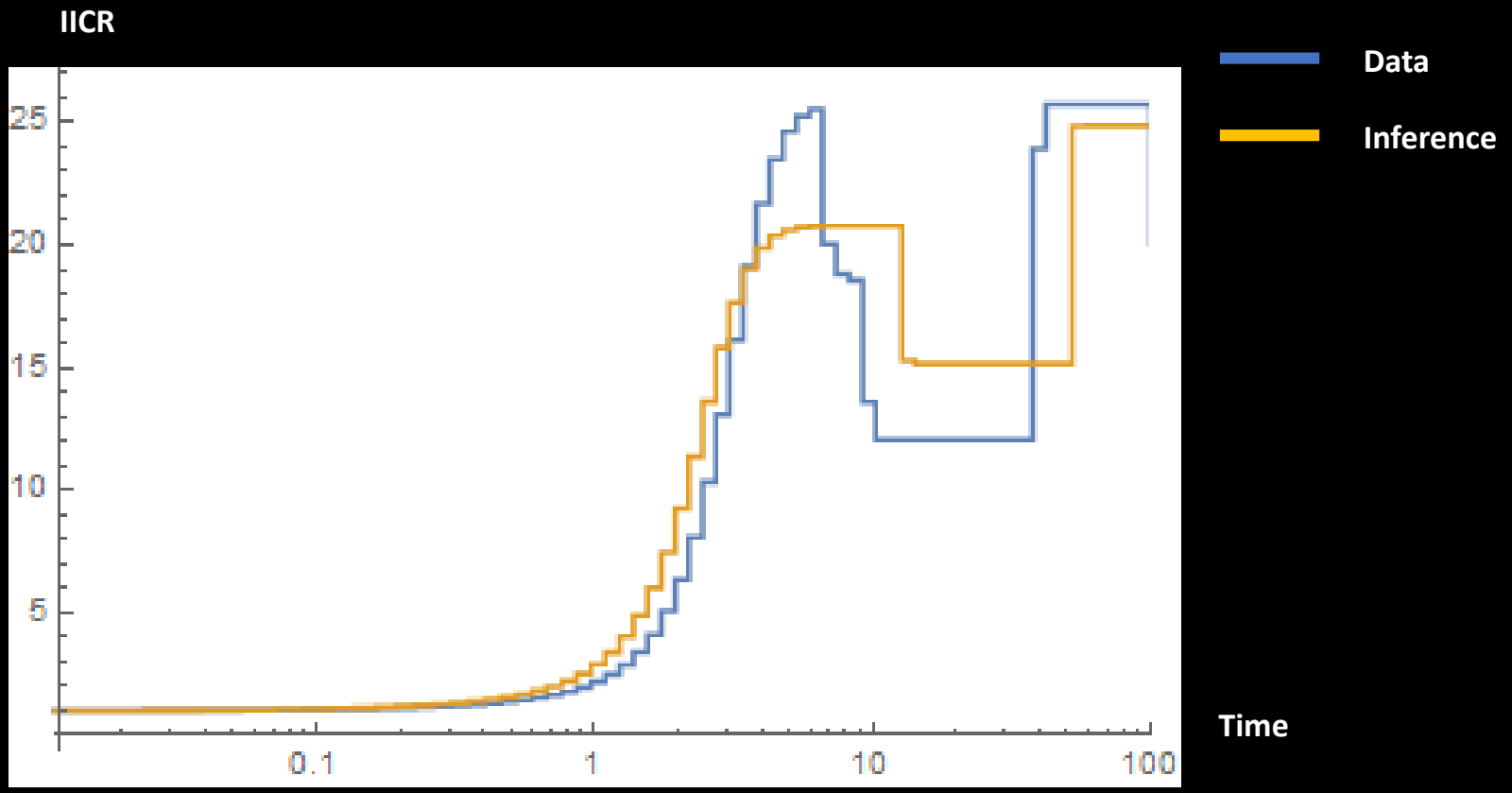

IICR
30
25
20
15
10
5
0.1
1
10
100
Time
Data
Inference

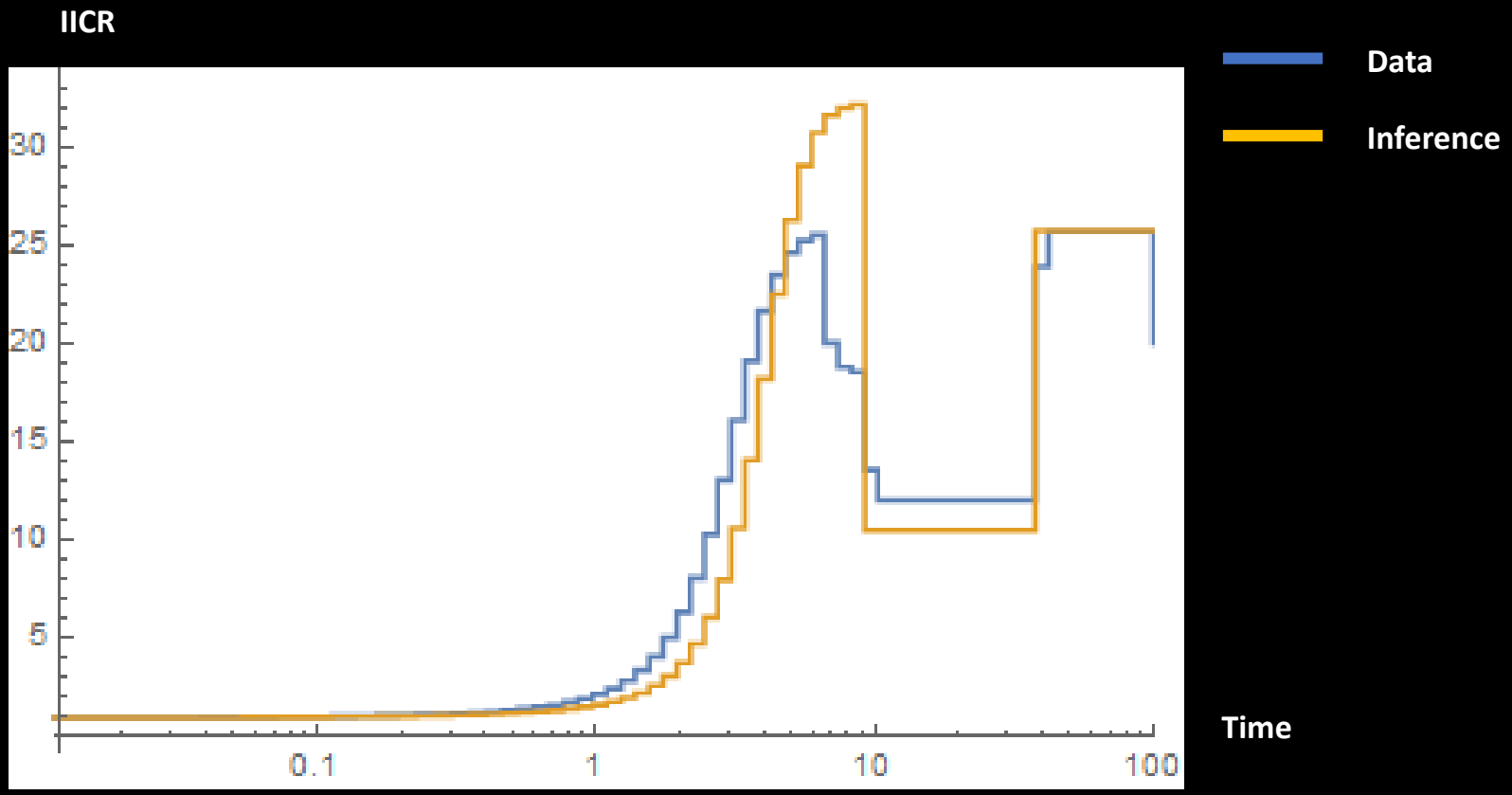

IICR
25
20
15
10
5
0.1
1
10
100
Data
Inference
Time

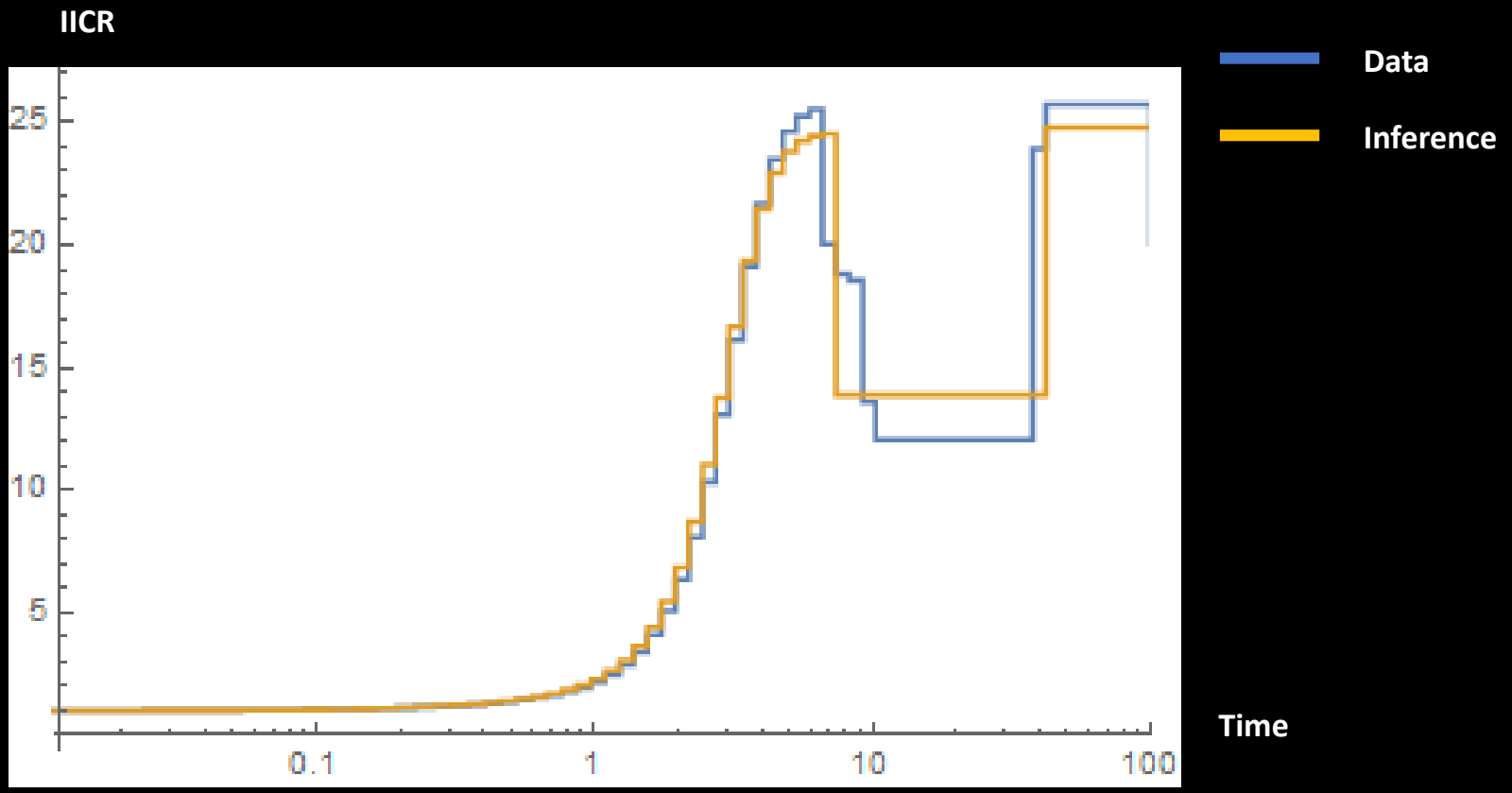

IICR
25
20
15
10
5
0.1
1
10
100
Data
Inference
Time

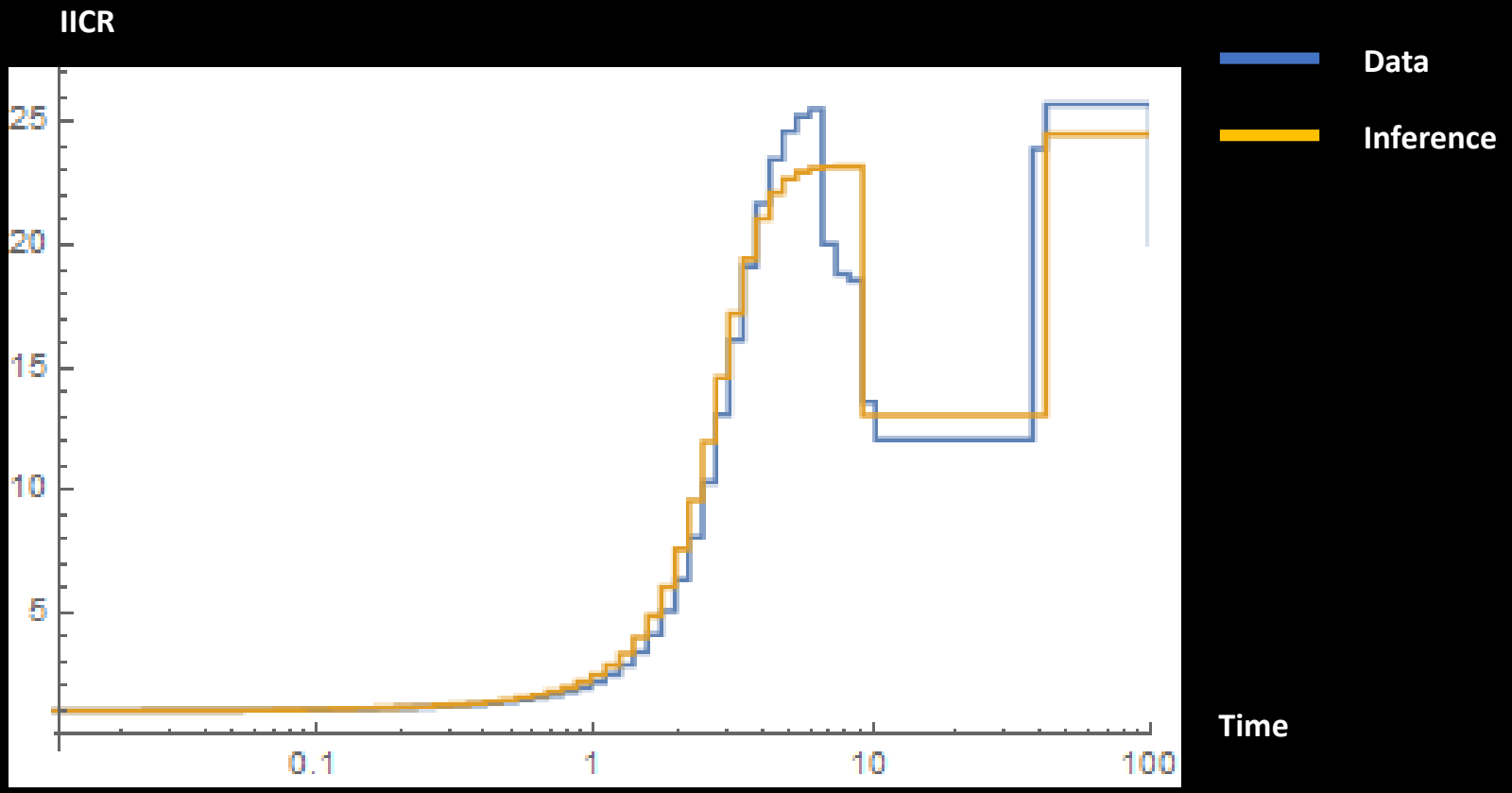

IICR
25
20
15
10
5
0.1
1
10
100
Time
Data
Inference

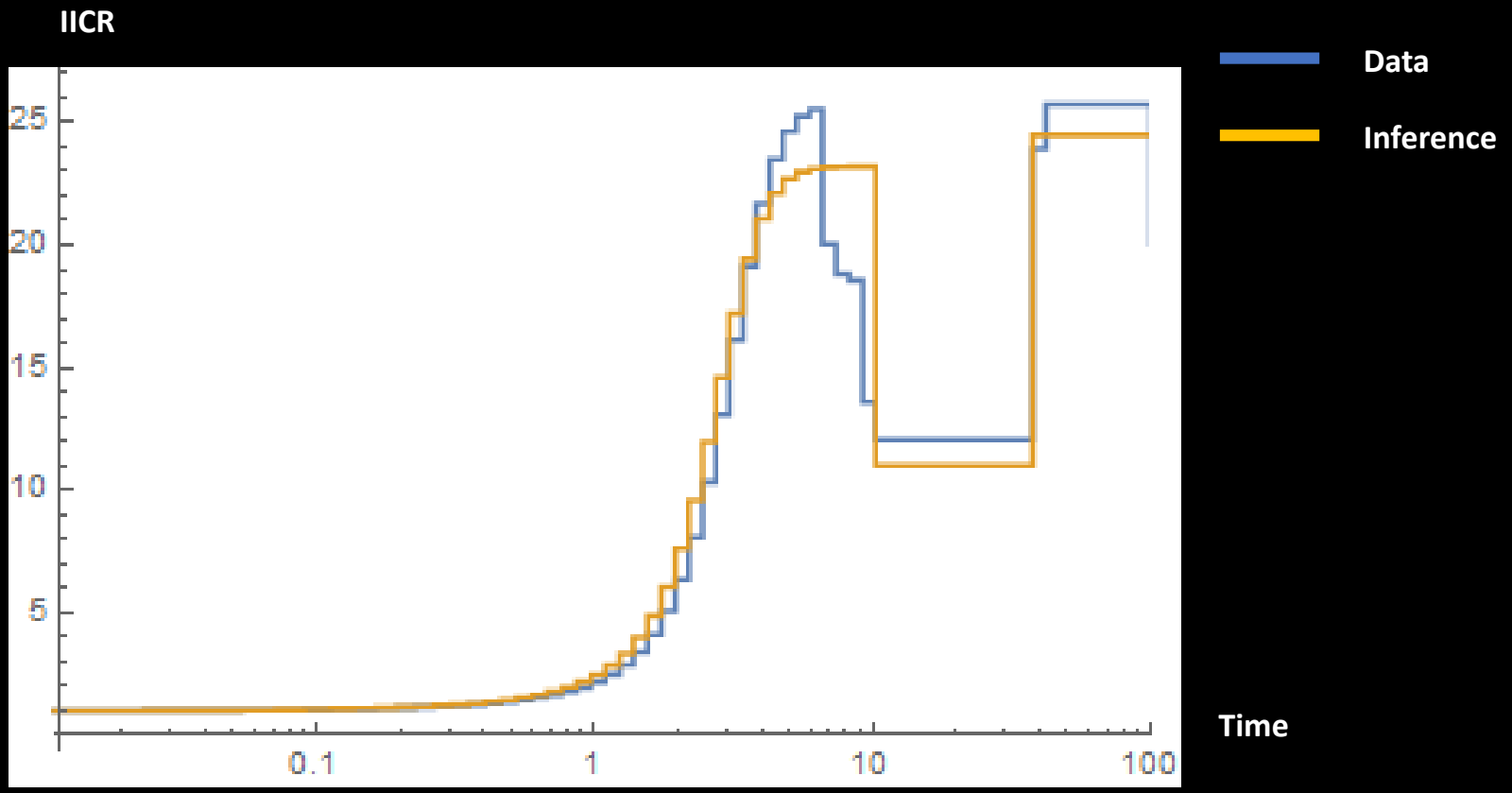

IICR
25
20
15
10
5
0.1
1
10
100
Data
Inference
Time

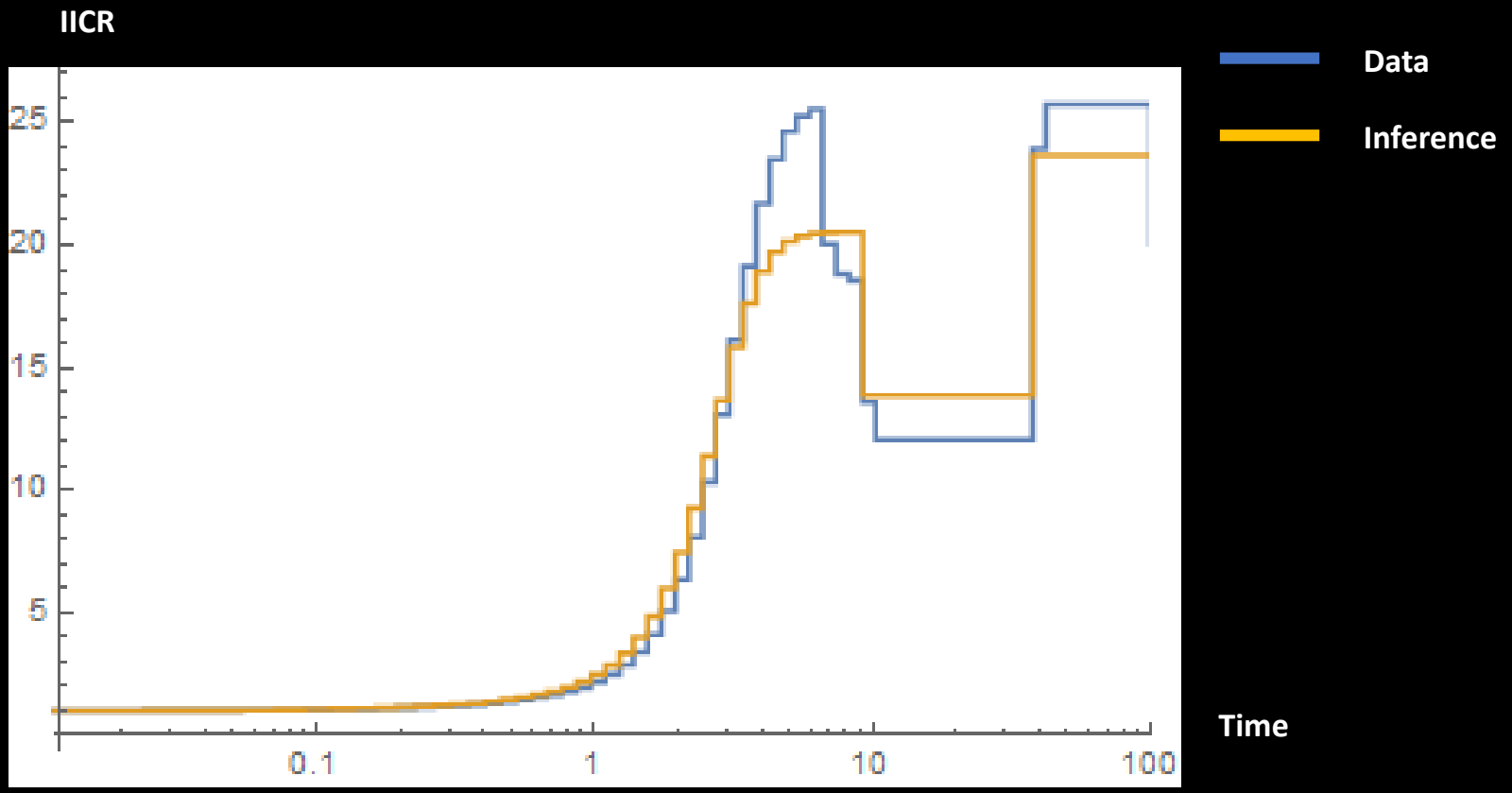

IICR
25
20
15
10
5
0.1
1
10
100
Data
Inference
Time

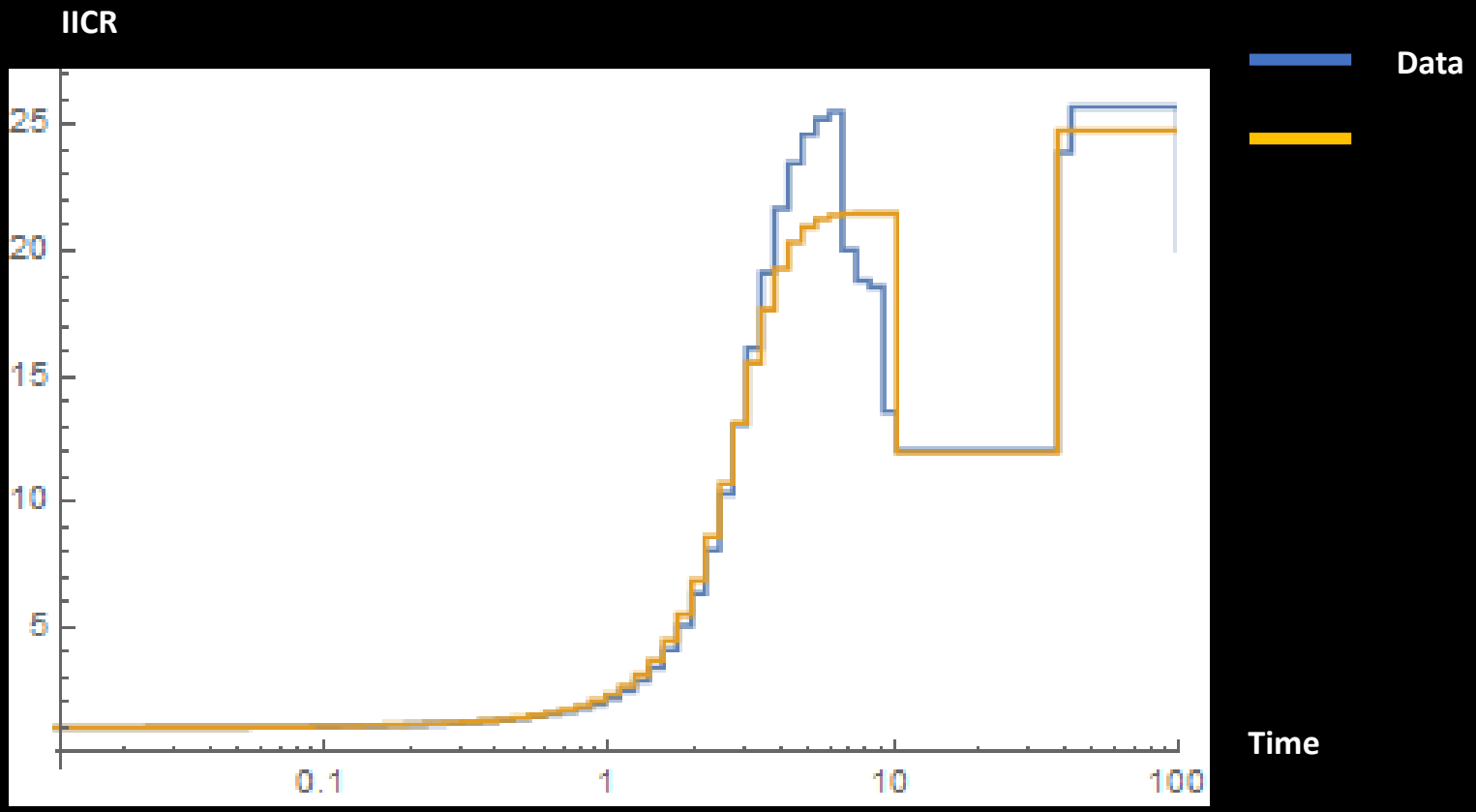

IICR
30
25
20
15
10
5
0.1
1
10
100
Time
Data
Inference

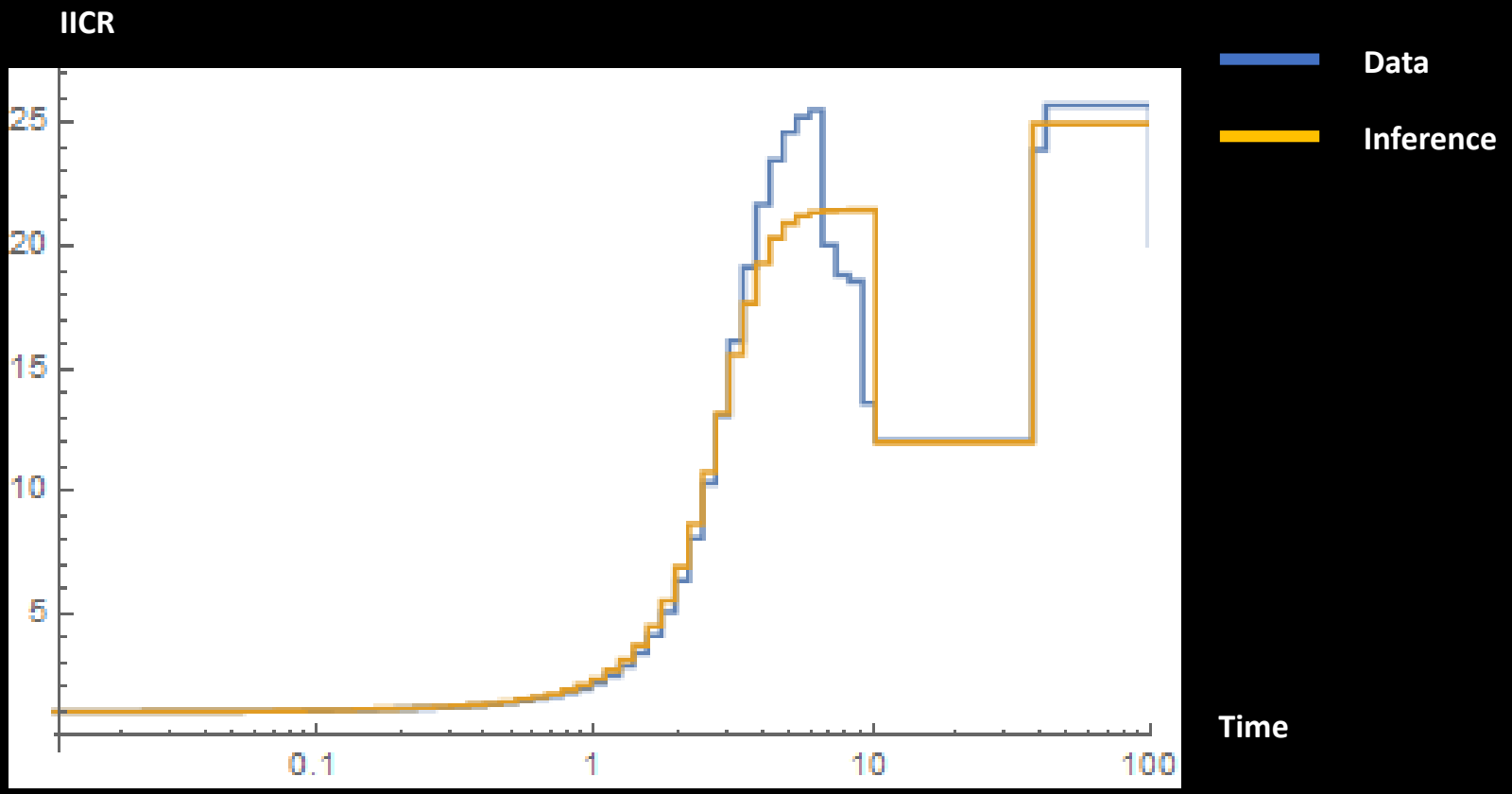

IICR
25
20
15
10
5
0.1
1
10
100
Data
Inference
Time

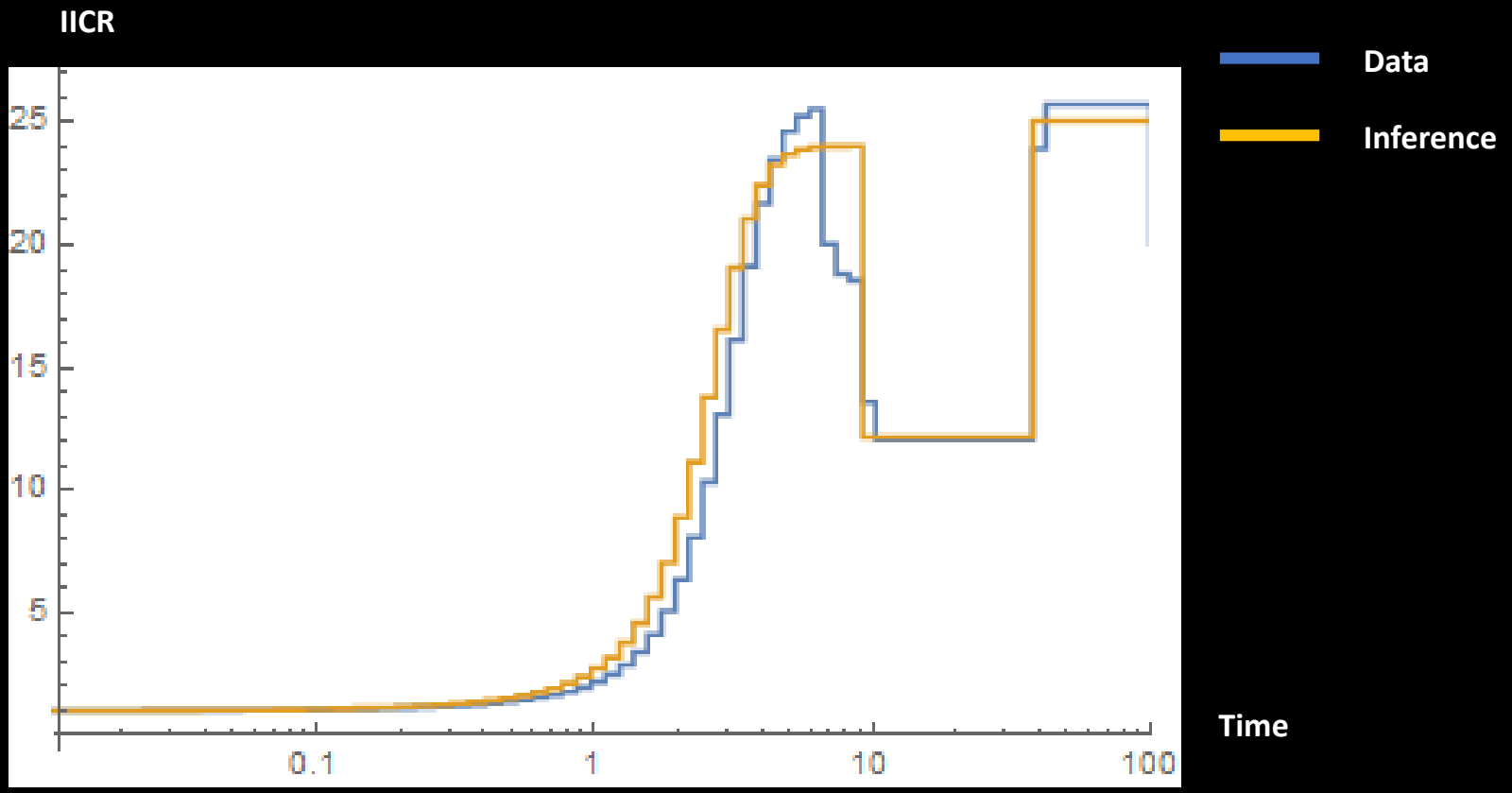

IICR
25
20
15
10
5
0.1
1
10
100
Time
Data
Inference

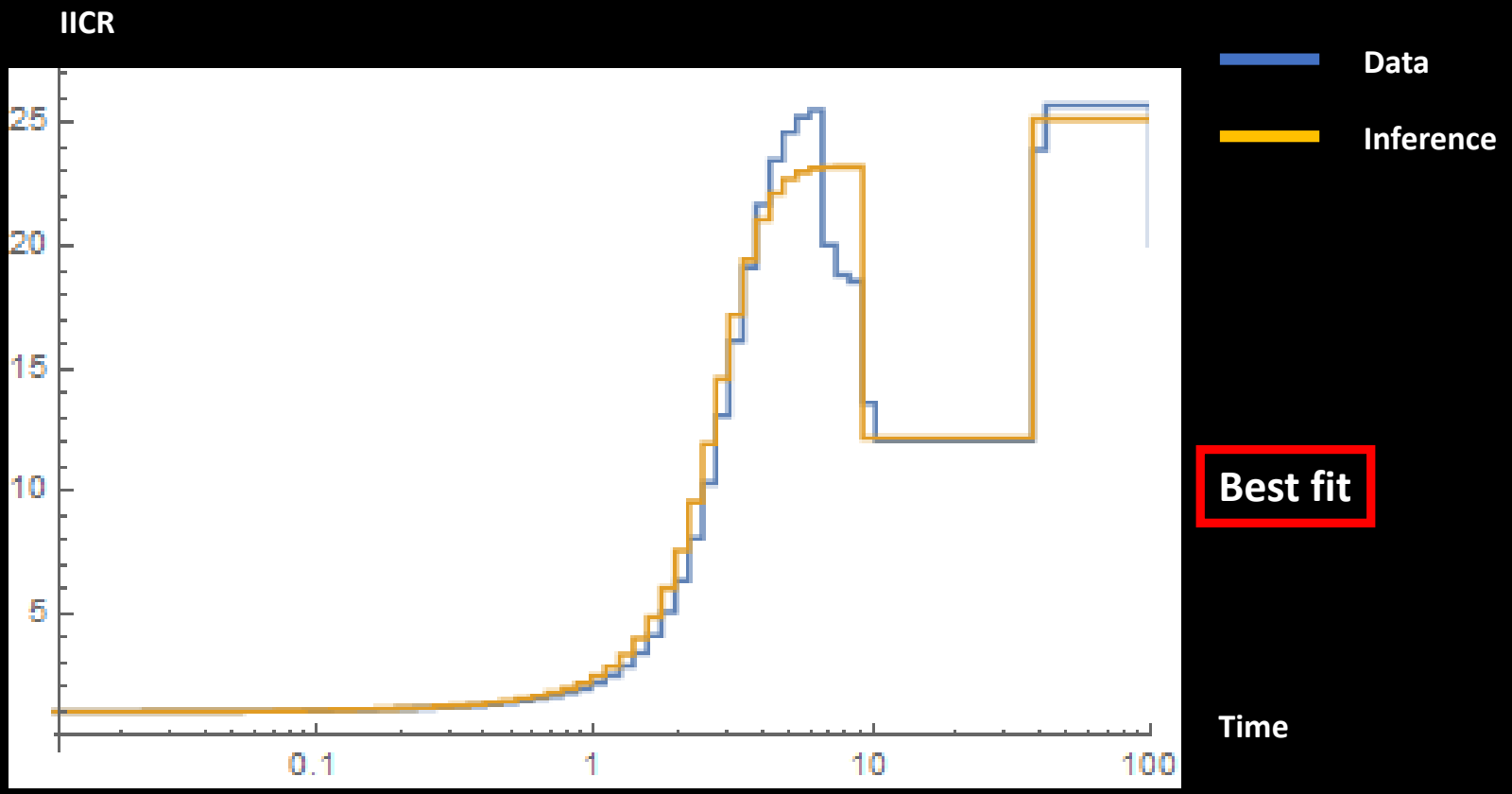

IICR
25
20
15
10
5
0.1
1
10
100
Data
Inference
Time

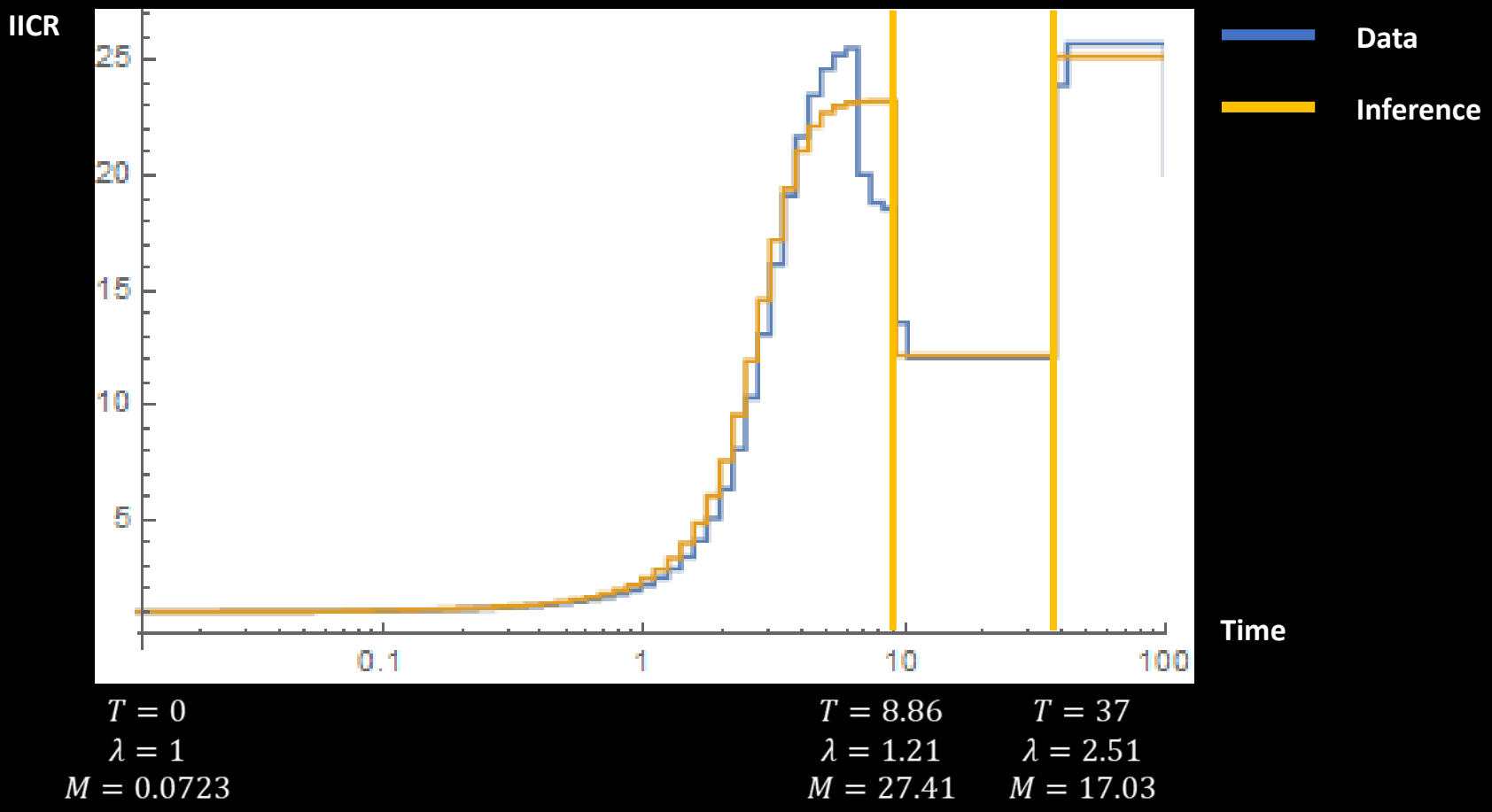

IICR
25
20
15
10
5
0.1
1
10
100
Data
Inference
Time


IICR
25
20
15
10
5
0.1
1
10
100
Data
Time

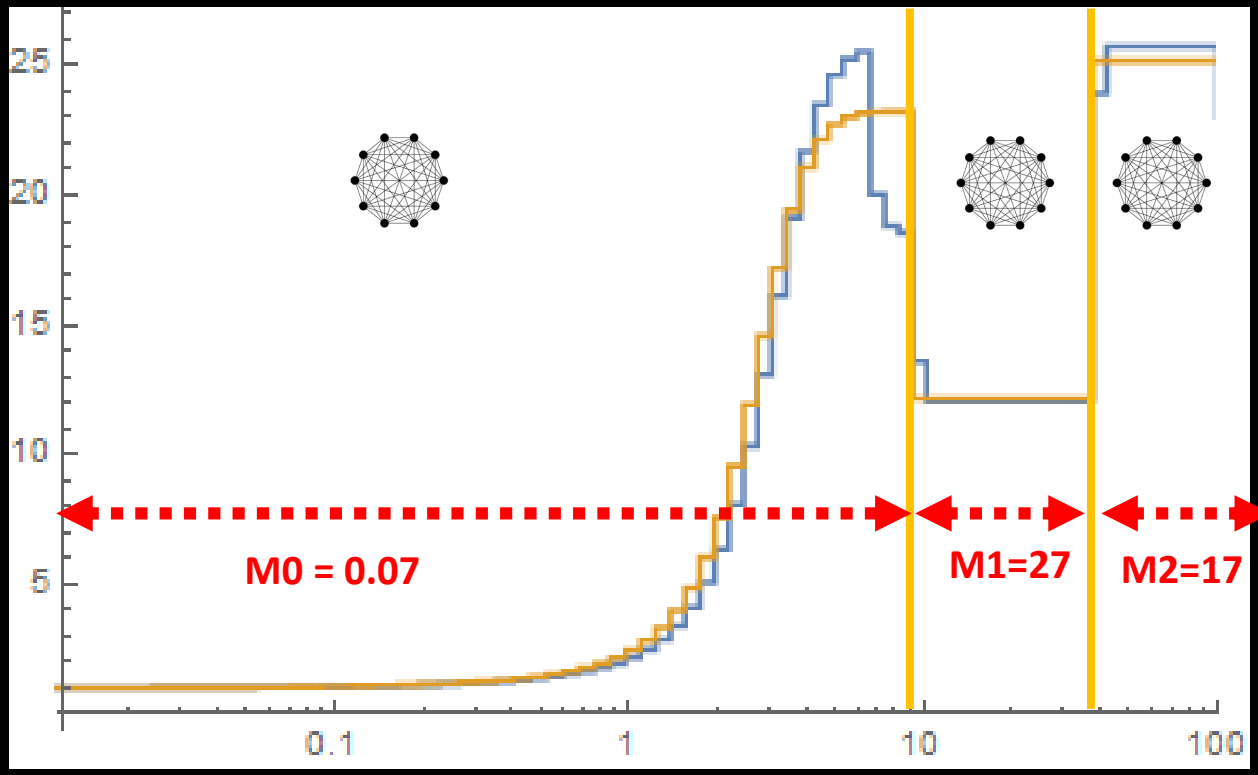

IICR
25
20
15
10
5
0.1
1
10
100
Data
Inference
Time


IICR
25
20
15
10
5
0.1
1
10
100
Time
Data
Inference


IICR
25
20
15
10
5
0.1
1
10
100
Data
Inference
Best fit
Time


IICR
25
20
15
10
5
0.1
1
10
100
Time
Data
Inference
$T = 0$
$\lambda = 1$
$M = 0.0723$
$T = 8.86$
$\lambda = 1.21$
$M = 27.41$
$T = 37$
$\lambda = 2.51$
$M = 17.03$

IICR



$T = 0$
$\lambda = 1$
$M = 0.0723$

$T = 8.86$
$\lambda = 1.21$
$M = 27.41$

$T = 37$
$\lambda = 2.51$
$M = 17.03$

t0 = 0

Time

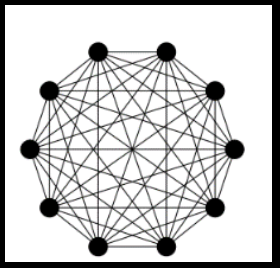
t1 = 8.86

t2 = 37

present

past

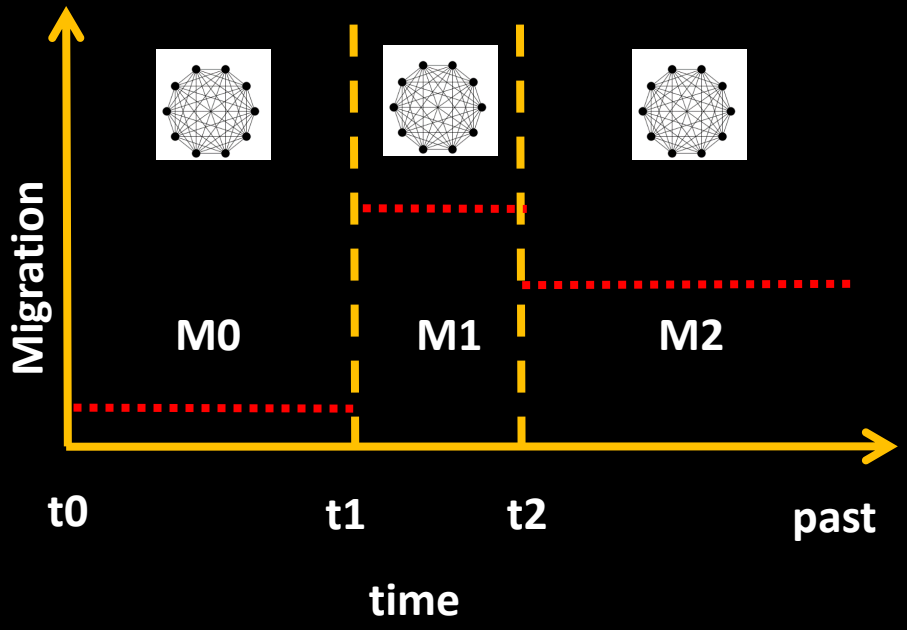
Application to simulated data



Connectivity Graph

n-island model

n: number of islands/populations
Nref: size of the islands/populations



Model with three components

Summary (1/2)

- The PSMC method uses the genome of a single individual and produces a wavy curve that is a function of time
- This wavy curve corresponds to what we call the IICR (inverse instantaneous coalescence rate)
- Under panmixia the IICR is equivalent to an effective population size (Ne) → Interpretation 1
- Under population structure the same IICR can be interpreted in terms of changes in connectivity (assuming constant population size) → Interpretation2
- It is possible to fit by hand model of population structure with a limited number of changes in connectivity

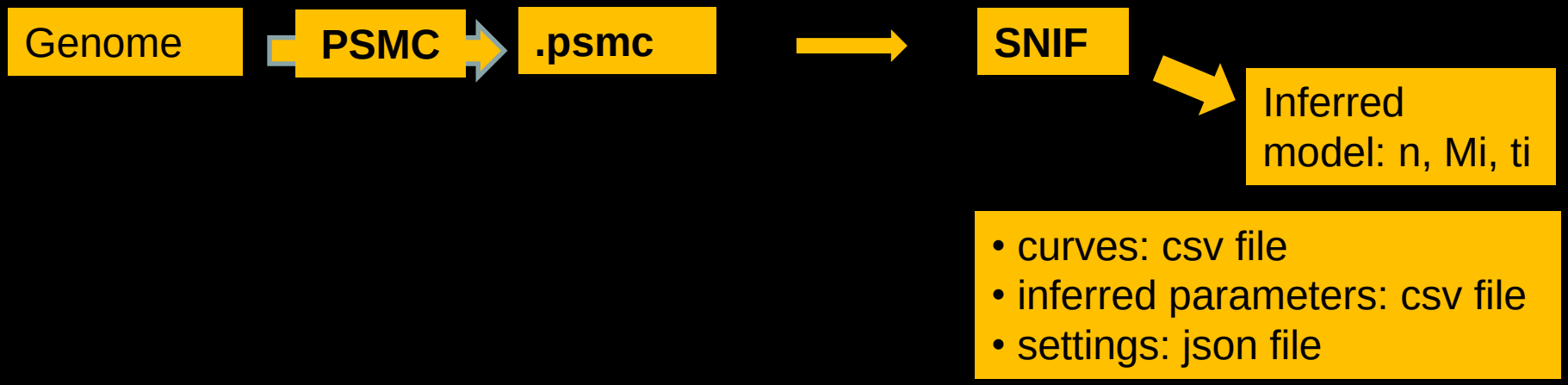
Summary (2/2)

- SNIF is a method that fits the best n-island model to a PSMC/IICR curve and infers the parameters of a non-stationary n-island model or "piecewise stationary n-island model".

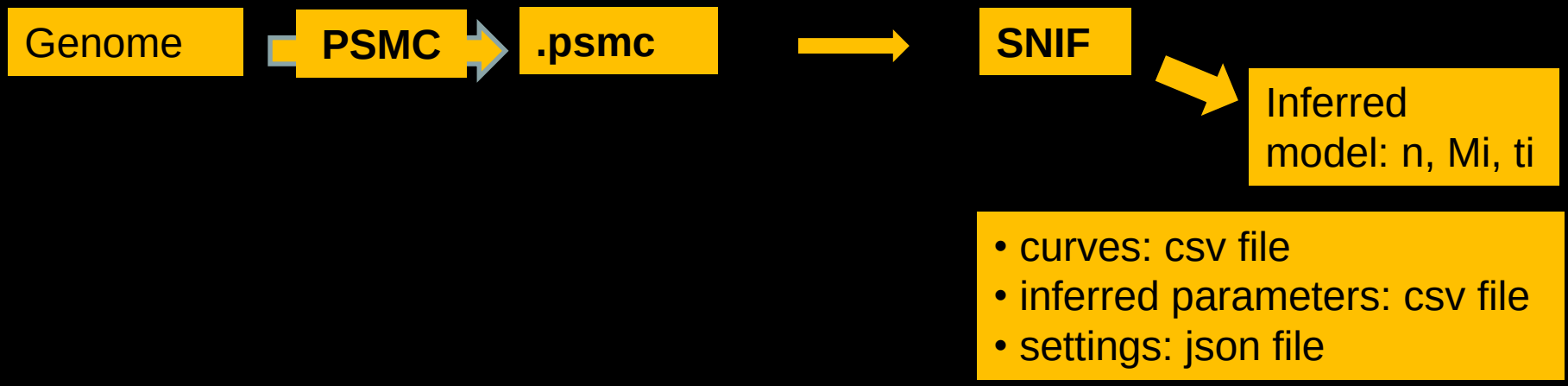
- We can infer (for a given number of components)
  - The number of populations (most methods assume that number and do not estimate it)
  - The size of the populations
  - The times at which gene glow (connectivity) changes
  - The values of migration rates (gene flow) for each of the components

- SNIF has been validated under very complex conditions using simulated data

SNIF


Genome
PSMC
.psmc
SNIF
Inferred model: n, Mi, ti
• curves: csv file
• inferred parameters: csv file
• settings: json file

SNIF



SNIF parameter files
- c: range nb of components
- n: range nb of islands
- w (omega): distance weight
- etc.

SNIF

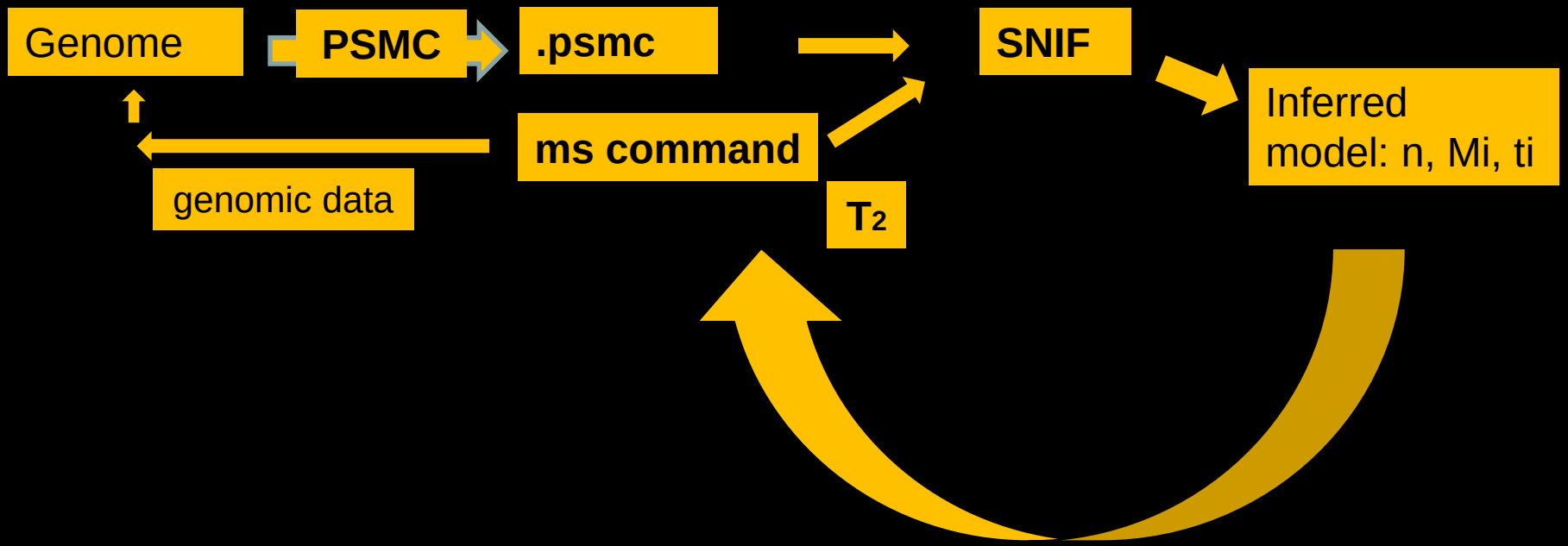
SHOULD YOU BELIEVE THE INFERRED MODEL?

VALIDATION ?

1. Simulate data under the inferred model
2. Use the simulated data as input for SNIF
3. Compare the results of this newly inferred model to the model used as input

SNIF parameter files
- c: range nb of components
- n: range nb of islands
- w (omega): distance weight
- etc.

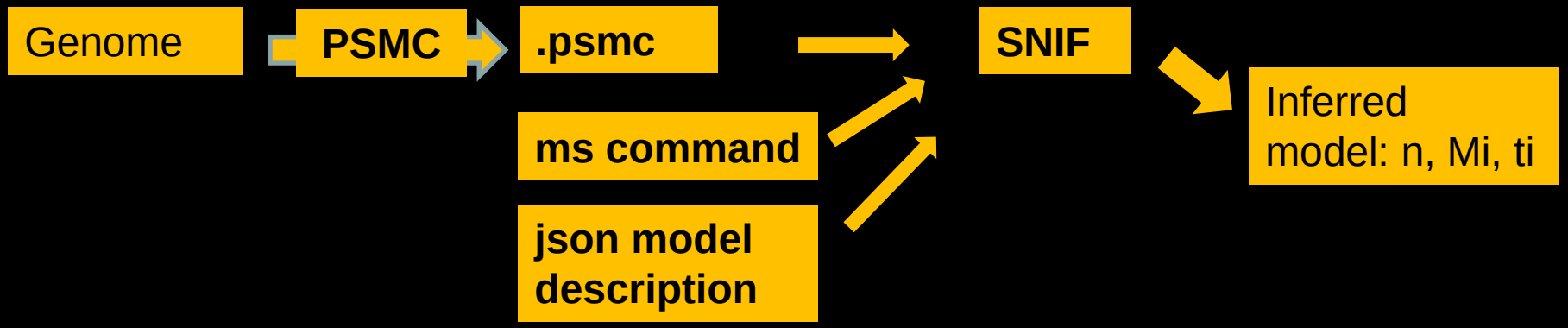
SNIF



SNIF parameter files

- c: range nb of components
- n: range nb of islands
- w (omega): distance weight
- etc.

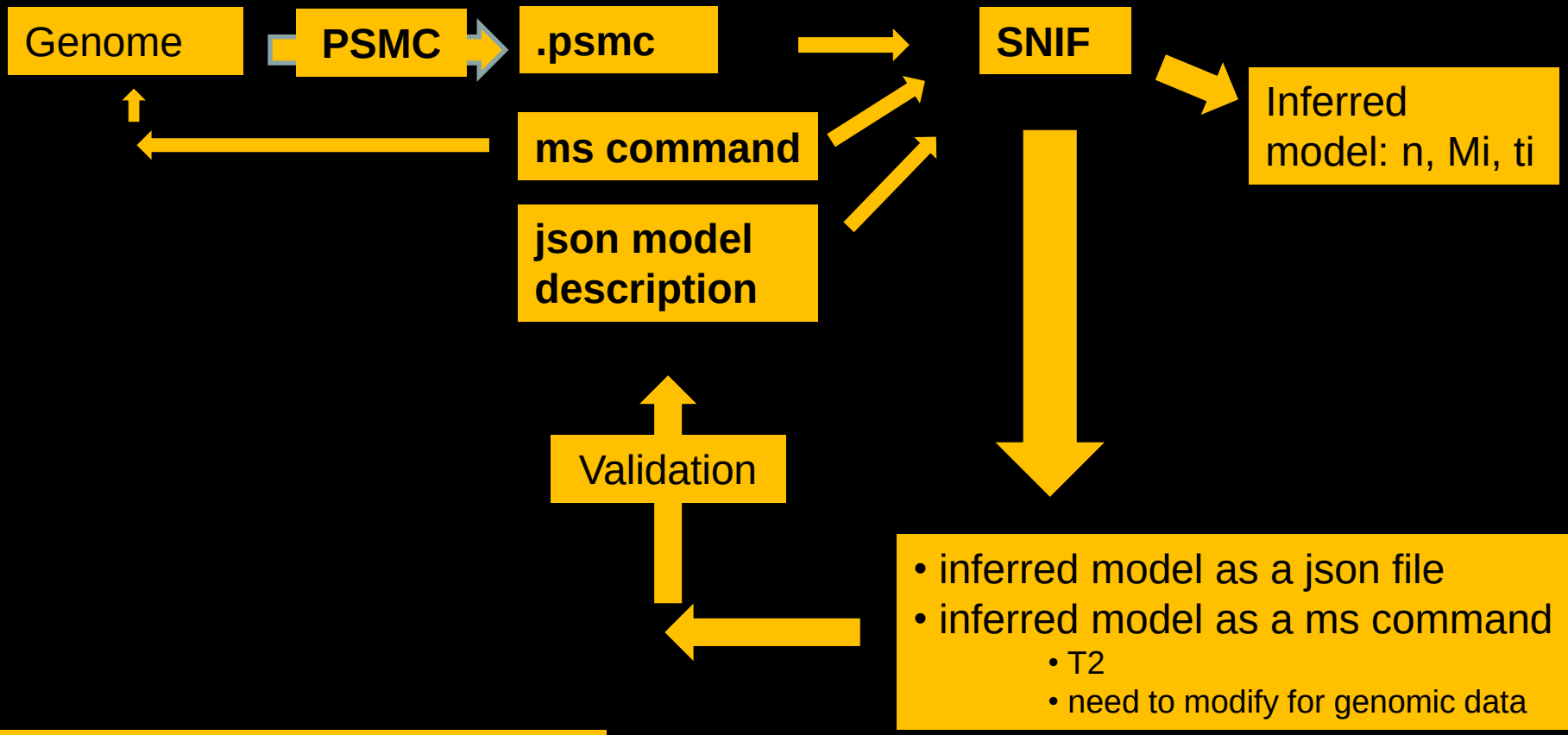
SNIF



SNIF parameter files

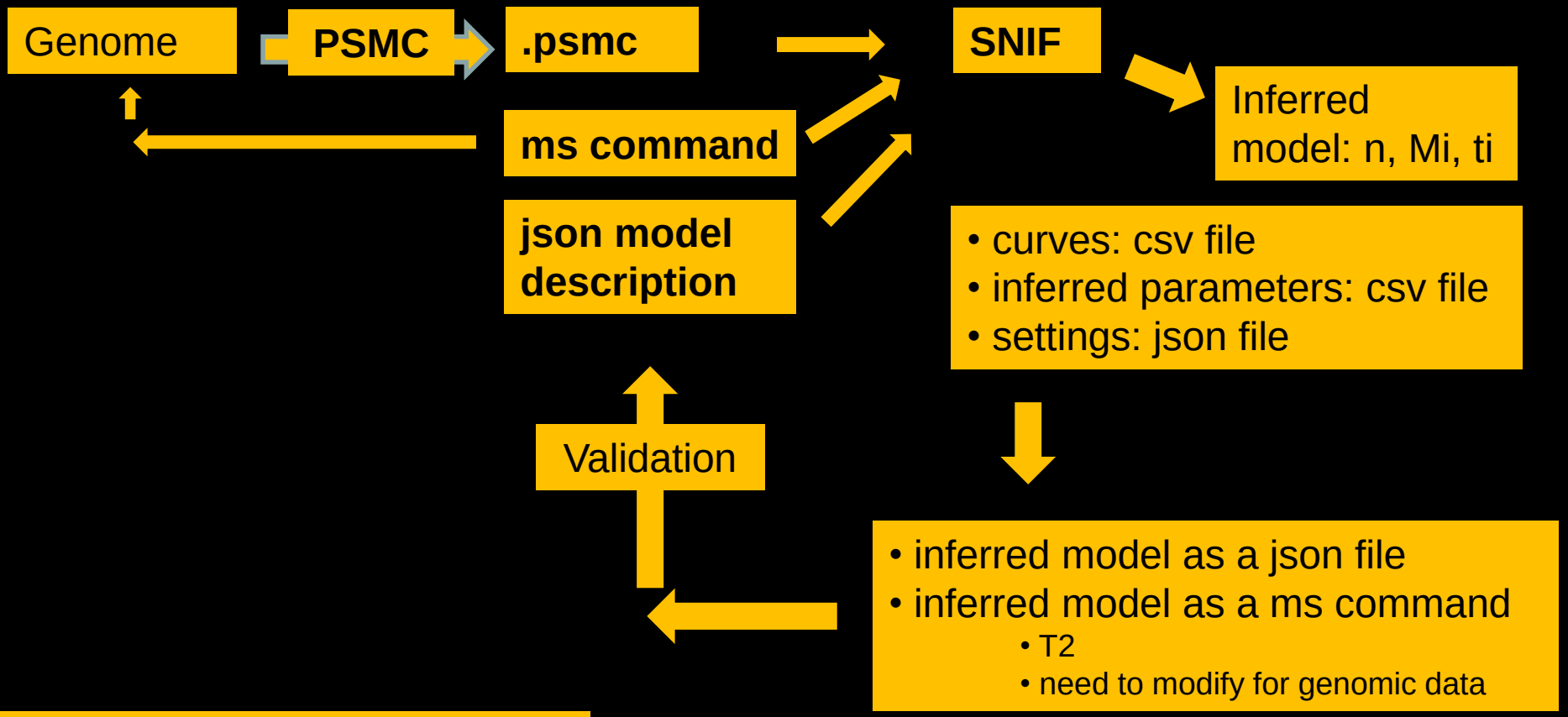
- c: range nb of components
- n: range nb of islands
- w (omega): distance weight
- etc.

SNIF



SNIF parameter files
- c: range nb of components
- n: range nb of islands
- w (omega): distance weight
- etc.

SNIF



SNIF parameter files

- c: range nb of components
- n: range nb of islands
- w (omega): distance weight
- etc.