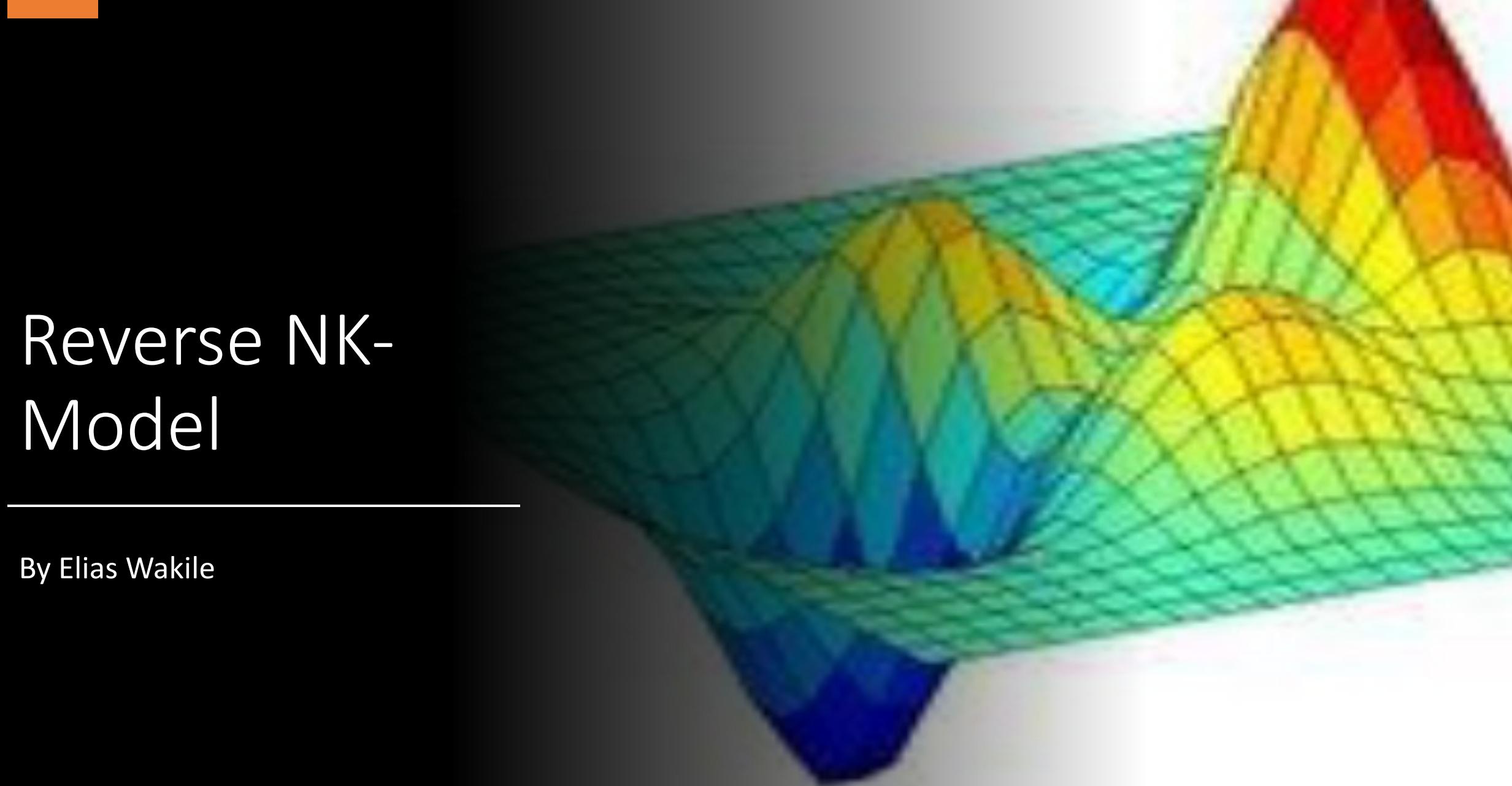


Reverse NK- Model

By Elias Wakile

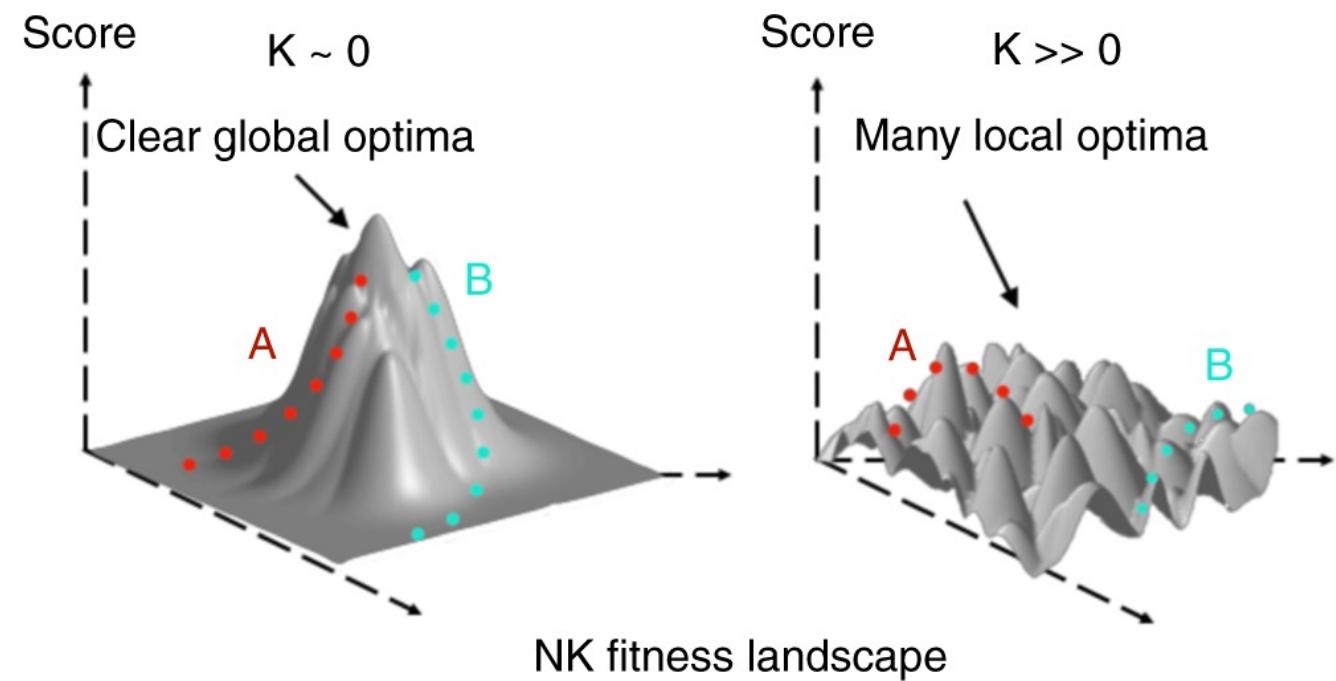


Introduction: Fitness & Epistasis

- The fitness of an allele describes the capacity of the allele to survive in nature.
- The Epistasis is the variation in the fitness of an allele constrained to the genetic background in which the allele is present, by means of “What is the contribution of an allele in a given genetic background?”.
- *The stochastic emergence of mutations may act as a significant diversifying force in evolution by causing the evolutionary divergence of isolated populations. [Mani & Clarke 1990, p. 29-37]. Additionally, the fixation of one allele might influence the fixation environment for other alleles, for example by the prevention of their occurrence. [Gillespie J.H. 1984, p.1116-1129] [Welch J. 2005, p.333-334].*

Reminder: The NK Model

- The multiple constraints we have previously enumerated might challenge us in the expression of the fitness of all the possible combinations of alleles. However, we have at our disposal a powerful computational method to express and study these genetic interactions, the random mode “NK Model” proposed by Kauffman and Levin [Kauffman & Levin 1987, p.11]*
- In the NK model, the objective is to determine the fitness function h given all the fitness functions of the subsequences (all possible combinations of K locus among the N we study) . Remember that a high K implies a rigid surface of the plot of the fitness function, while a lower K implies a smoother surface. We simply sum the fitness of all the allele combinations in our genome



What is the Reverse NK-Model ?

So what is the Reverse NK-Model ? It is the model, that computes the fitness functions of all K subsequences of our N locus, given the genome of a population Ξ the respective fitness of the individual having a certain genome h , and an epistasis parameter K . But how are we going to choose our parameter K ?!



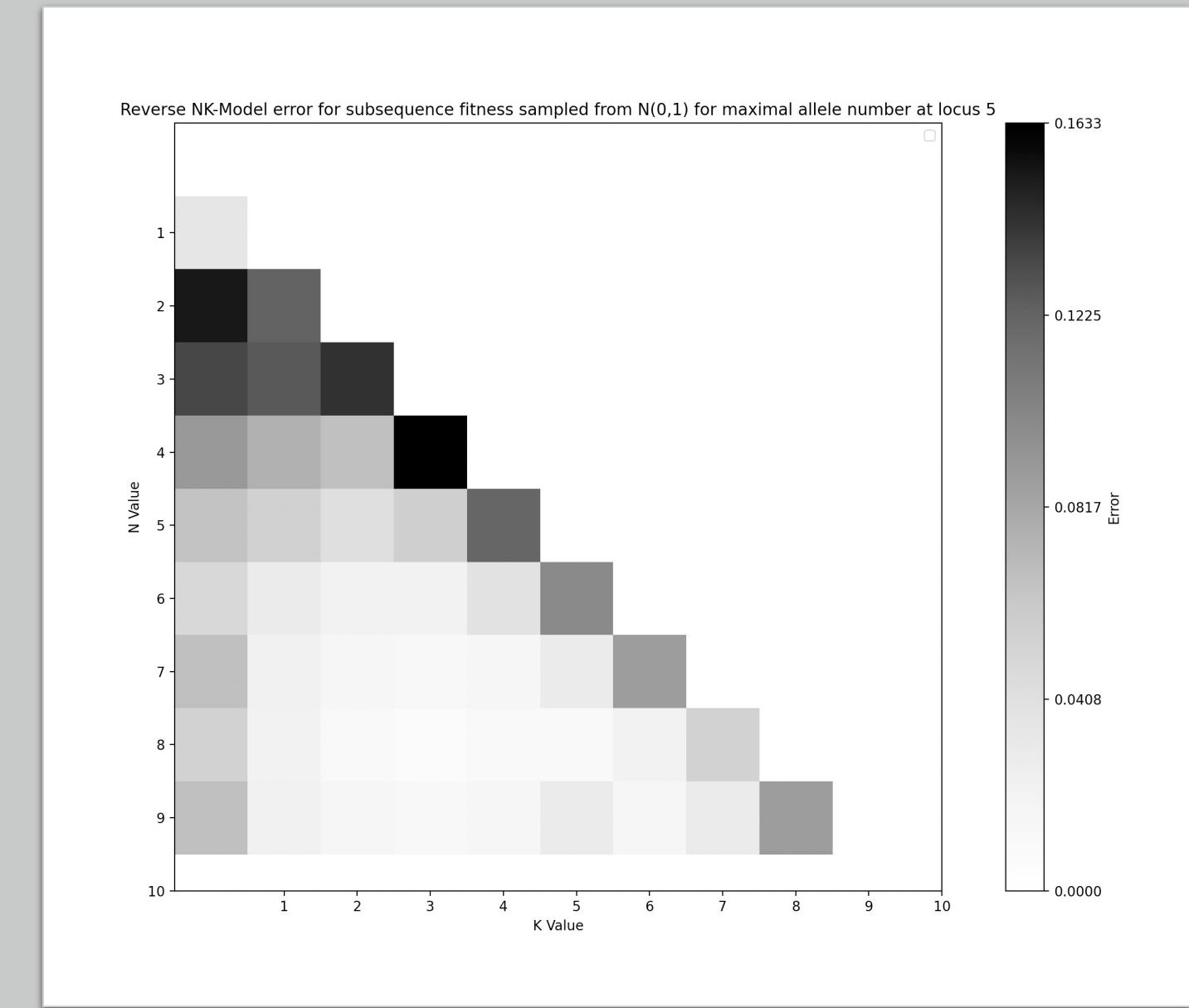
The K parameter does not pose much of a problem, since it was already demonstrated by previous researches that the fitness equilibrium reached during simulations was barely affected by the epistasis parameter K. [Ohta T. 1998, P.83-90]. Consequently, we get the advantage of tuning the epistasis parameter in order to obtain a richer model, but, at the cost of computational ressources, especially runtime.

How does the Reverse NK-Model work?

- First, we hash the genomes of the population into an index matrix X denoting what combination of alleles at what locus while each row represents an individual genome and each row represents a combination of K alleles at different locus
- Assuming we only have the fitness of each individual genome, we fill those into a vector y such that each of his row will have the respective fitness of the hashed Genome in the X matrix.
- We will compute the singular value decomposition of $X = U \cdot \Sigma \cdot V^T$. U is the orthogonal matrix of the left eigenvectors of X , Σ is a diagonal matrix having at its diagonal the respective eigenvalues of X , V is the orthogonal matrix of the right eigenvectors of X .
- Finally, we will find the vector w that realizes $Xw=y$ by the Moore Penrose pseudo-inverse method $w = V \cdot \Sigma^{-1} \cdot U^T \cdot y$. If XX^T is invertible our solution is unique, otherwise it will find us an optimal solution for the problem. Therefore, it is guaranteed we will always have a solution. By w we are able to deduce the fitness of every combination.

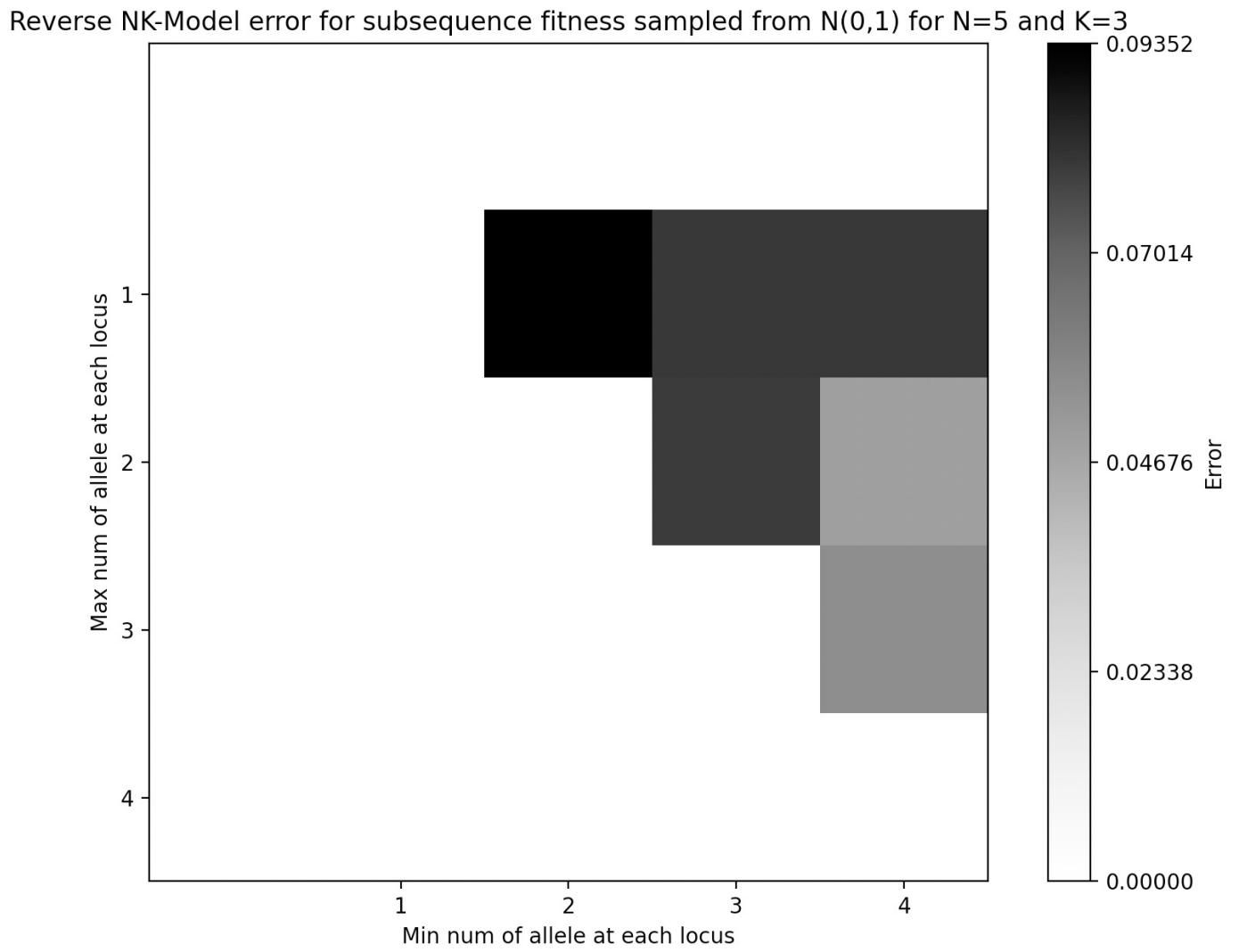
How does the Reverse NK-Model work?

- The algorithm runs in polynomial time and polynomial space if we regard K as a constant!
 - We tested our algorithm for different parameters, it generally yielded a very loss (error) converging to 0. Why did we measure the error and not the accuracy ? Because the Moore Penrose method solves an Empirical Risk Minimization optimization problem.

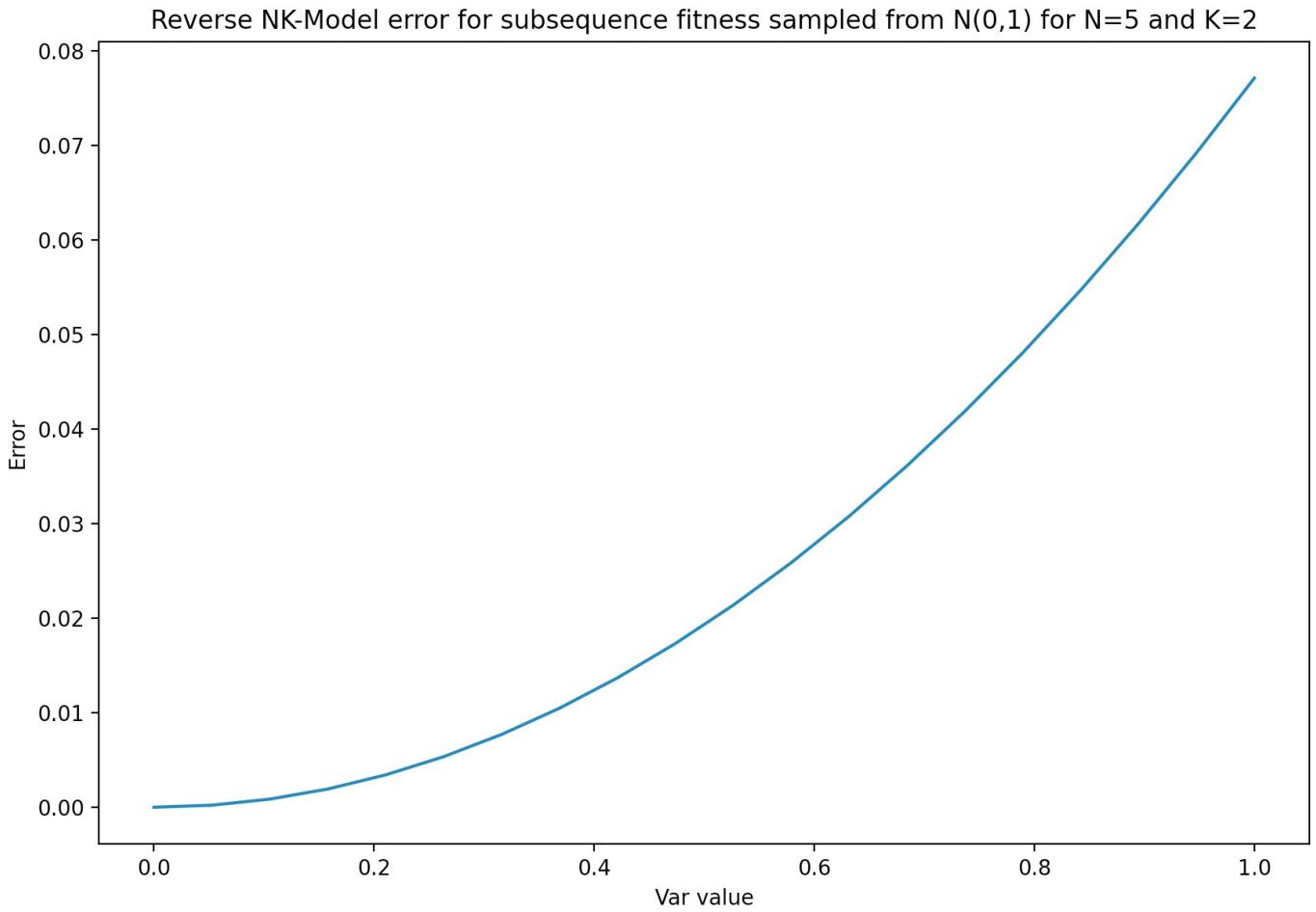


More Results

?



More
Results!



Bibliography

- Gillespie, John H. "Molecular evolution over the mutational landscape." *Evolution* (1984): 1116-1129.
- Kauffman, Stuart, and Simon Levin. "Towards a general theory of adaptive walks on rugged landscapes." *Journal of theoretical Biology* 128, no. 1 (1987): 11-45.
- MacAusland, Ross. "The Moore-Penrose inverse and least squares." *Math 420: Advanced Topics in Linear Algebra* (2014): 1-10.
- Ohta, Tomoko. "Evolution by nearly-neutral mutations." In *Mutation and Evolution*, pp. 83-90. Springer, Dordrecht, 1998.
- Van Loan, Charles F., and G. Golub. "Matrix computations (Johns Hopkins studies in mathematical sciences)." *Matrix Computations* (1996).
- Welch, John J., and David Waxman. "The nk model and population genetics." *Journal of theoretical biology* 234, no. 3 (2005): 329-340.
- Wright, Sewall. "The roles of mutation, inbreeding, crossbreeding, and selection in evolution." (1932): 355-366.



Thank you for listening!
Any questions?