Nonlinear Dynamics 101:
The NK Model
Kevin Dooley, Arizona State University

Evolutionary models enable researchers to examine basic questions about the dynamics of an adaptive population. Such models tend to be generic, in that the same model can be applied across a wide variety of types of populations, and yet still yield valid and insightful results. Most models of evolution are based on the concepts of *variation* and *selection*. In Darwinian models of evolution, variation occurs through random mutation or recombination, and selection occurs through the ability of the entity to survive so as to pass on its genetic code to its progeny, manifested through recombination. In more general terms, variation occurs through a change in the characteristic traits of the entity, and selection occurs via the fitness of the new structure.

For example, consider a genotype, or genetic string, of length N=5: {X1 X2 X3 X4 X5}. The overall fitness of the genotype depends on the state of each of the five genes X1-X5, and possibly their interactions. For example, if the genotype represented traits of a cat: {speed, hairiness, hearing ability, visual ability, curiosity}, then the overall fitness of the cat is determined by whether the cat is slow or fast, bald or furry, deaf or hearing-sensitive, blind or visually-sensitive, and curious or not. The states of each gene (i.e. slow, fast) are called alleles.

There also may be interactions; a blind cat might be better off not being curious, but a visually able cat might be better off being curious. The extent to which any gene's fitness contribution depends on the allele or state of another gene can be defined as another parameter, K, which denotes the amount of interconnectedness in the system. A system with K=0 corresponds to one where each gene contributes independently to fitness. A system with K=N-1 corresponds to one where each gene's fitness contribution depends on all other genes in the genotype.

As a practical example, consider the genotype defining the design of a bicycle. The overall fitness, or quality of the bicycle is determined by the quality of its parts: handlebars, gears, wheels, frame, etc. As bicycles are made purposefully modular, the bicycle is a system with K=0—you can select different tires (within reason), and that will not impact the fitness contribution due to the quality of the bicycle frame. At the opposite end, consider the composition of a pharmaceutical drug, which is essentially a chemical recipe. One cannot simply change the amount of one chemical constituent without changing many or all of the others—the fitness of the drug depends on the constituents in an integral way. A drug would be an example of a system with high K.

The *NK model* (defined by the two parameters, N and K) captures these elements of the evolutionary process; it has been studied most extensively by Stuart Kauffman, whose grand summary of work is presented in *Origins of Order* (Kauffman, 1993). It provides a structure by which to simulate vast populations of genotypes evolving over all different sorts of systems and environments (configurations of N and K). Kauffman argues that the NK model stands in contrast to neo-Darwinian interpretations of evolution, and instead points out the important role of self-organization: "Ordered properties can persist even without selection, or under weak selection…Adapting systems exhibit order not because of selection but in spite of it" (p. 35).

As a researcher, one can use the NK model in two ways. The first step is to model the system we have in mind in an NK framework. Specifically, we need to consider what the traits of the system/entity are that would contribute to fitness, and how they interact. This gives us the N and K parameters. Then we can either:

1. Take the results that Kauffman obtained from his simulations of the generic system and impose them upon our system—this is primarily an act of translation.

2. Use the architecture of the NK model as a starting point, and then add to it, relative to the specific nature of the system/entity in question. We can then simulate this modified NK model to observe patterns and draw conclusions.

The NK model enables the following research questions to be answered:

1. How many local optima exist?
2. What is the average number of improvement steps on an adaptive walk to local optima?
3. What is the average number of mutations in such a walk?
4. What is the ratio of mutations tried to mutations accepted?
5. What is the average, maximum number of local optima?
6. How do improvement rates dwindle?
7. How many genotypes can climb to the same local optima?

Here are some general findings that emerge from Kauffman's simulation of the NK model. In these simulations, a population of genotypes is allowed to evolve towards configurations of greater fitness. A genotype may change one or more of its genes at any given time, in a random fashion. If the change is beneficial, then the genotype "moves", i.e. one or more of its genes change. Both mutation and recombination can be mimicked in the model. The actual assignment of fitness values to particular gene alleles is random, although it turns out results are fairly robust to this assumption. The collection of all of the possible genotype configurations and their corresponding fitness values is referred to as a *landscape*. Thus evolution of the genotype can be considered adaptive walk on a genetic landscape. These points are based on my own reading notes from *Origins of Order*:

1. When K=0
   - Each genotype lies on a connected path to the single optimal genotype.
   - The landscape is maximally smooth and the "adjacent" points (genotypes) are highly correlated.
   - The expected number of steps to optimum (using a one-mutant neighbor) is N/2.
2. When K=N-1 (46)
   - The landscape is completely uncorrelated
   - The number of local fitness optima is extremely large.
   - The expected fraction of fitter one-mutant variants dwindles by ½ each adaptive step.
   - The length of adaptive walks is very short.
   - The ratio of accepted to tried mutations scales as $\ln\{N\}/N$.
   - A genotype can only climb to a small number of local optima.
   - Only a small fraction of the genotypes can climb to any given optimum.
   - As N increases, the fitness of local optima falls toward the mean fitness of the entire space.
3. As N increases, a *complexity catastrophe* occurs. Each gene contributes less to overall fitness, thus a disadvantageous mutation may not be selected against, and advantageous ones will not contribute as strongly. Also, the number of conflicting constraints increases, thus causing poor compromises. This cannot be overcome by selection—selection can only work on what is it is given.
4. Various K
   - Most aspects of landscapes are insensitive to any parameter except N and K.
   - If K is fixed while N increases, a complexity catastrophe does not occur. It occurs when K increases with N.
   - The fitness of optima for small values of K (2,4,8) is higher than for K=0: "Low levels of epistatic interaction appear to buckle up the landscape like heaving mountain ranges" (p. 56).
   - As the genotype veers away from local optima, the drop is likely to be more precipitous as K increases.
   - For small values of K and two alleles (important distinction—this is not true for "continuous" alleles.), the highest optima tend to be near one another.
   - The highest peaks drain from the largest basins.
5. A "long jump", one changing more than one gene at a time, if beyond the correlation length of the landscape, is essentially a jump on an uncorrelated landscape. Thus waiting times double

after each fitter variant is found. Long jumps suffer from complexity catastrophes. Thus to maintain the highest fitness, the landscape should be low K, and searches should proceed locally.

6. There are three natural time scales in searches:
   - Early in the adapting process, local solutions are low in fitness because of design constraints, whereas long jumps may lead to much better solutions because they are uncorrelated with initial conditions. Most members of the evolving population will end up here.
   - In the midterm, local searches yield better results, and thus proceed.
   - Once at a local optimum, a long jump is needed before improvement can be found, but many such long jumps may be unsuccessful.

7. Since much of the radiation (divergence, speciation) leads to genotypes with poor fitness, there will tend to be extinctions. Speciation and extinction may be separated by long periods of stasis—punctuated equilibrium.

8. Genes with differential K:
   - K low—low entrenchment; K high—high entrenchment
   - When fitness is low, genes with high entrenchment are more likely to change; as fitness increases, genes with low entrenchment change.
   - If the environment is fixed or slow, then low entrenchment genes will dominate most change. If the environment is changing, then high entrenchment genes will dominate change. If the environment changes rarely but dramatically, then high K genes will change first, followed by low K genes.

9. Convergence and divergence are natural dynamics of adaptive walks: "Convergence can represent adaptation climbing from different initial points to either the same or nearby peaks in the fitness landscape." (p. 85). Relative rates of convergence and divergence on environmental volatility. When landscapes change drastically, fit types are cast back to low fitness. This leads to a burst of radiation, because many paths lead up. Highly entrenched traits will change but then become locked in.

10. *Error catastrophe*: On smooth landscapes, after evolving to local optima, rates of mutation carry genotypes down into the valleys. If the rate of mutation is greater than the rate of climb (which is low in a smooth landscape), then types will slip to lower fitness and be trapped there. Conversely, on a rugged landscape, optima tend to be surrounded by deep chasms, thus selection forces are strong and types tend to not fall off peaks due to mutation.

11. On adaptation: "Evolvability, the capacity to search a reasonable fraction of the space, may be optimized when landscape structure, mutation rate, and population size are adjusted so that populations just begin to melt from local regions of space…. Sustained fitness may be optimized when landscape structure is tuned so that the sides of fitness peaks are steep enough to offset mutation rates and the rate at which the landscape is deformed by abiotic or coevolutionary forces" (p. 95)

12. Recombination is expensive, but ubiquitous. Therefore it must have some value as a process of evolution. Recombination is useless on uncorrelated landscapes. It is useful when (a) high peaks are near one another and thus carry mutual information about their fitness, (b) parts of the system are quasi-independent, so parts of the system can be interchanged with some chance that the new type will get the "best of both". In K>0 situations, because peaks tend to be near one another, pairs of peaks (as represented in recombination) contain mutual information (as realized in redundant or identical gene sequences) about where good genotypes are located.

13. As mutation rates increase, divergence increases; initial condition effects decrease.

Now, let me give an example application to show how I have used the model to develop propositions about a system of interest (Choi and Dooley, 2000). Let the genotype represent a "supply field", defined as the set of all firms that could be within a focal firm's supply base—so it includes current suppliers, as well as those that potentially could be suppliers. Each gene represents a single firm, and a "1" in that gene corresponds to that firm being within the actual supplier base of the focal firm, and a "0"

corresponds to it not being a supplier. The genotype thus defines who is in and who is out of the focal firm's supply base, and the supply field's fitness depends on the quality, cost, and timeliness of the product being manufactured. Adaptation of this genotype corresponds to the common problem of *supplier selection*—whether to produce something in-house or not, and if contracted out, which supplier(s) to choose for which components of the product.

The NK model would tell me that if the focal firm's product can be made modular, such that the fitness of the overall product only depends independently on the fitness contribution of each supplier's part, then supplier selections can be made independently, and incremental improvement, focused on each separate supplied part, will be most advantageous. It also indicates that recombination would be useful—recombination in this context means learning from other focal firm's configurations via benchmarking and reverse engineering. Conversely, if the focal firm's product is integral, such that that the fitness of the overall product depends interdependently on the fitness contribution of each supplier's part, then incremental improvement, focused on each separate supplied part, will be disadvantageous. Rather, improvements are more likely to occur through more radical trial and error learning, and it may be necessary to employ "multiple search parties", i.e. have parallel developments, to ensure that the final configuration is not only locally optimal.

In summary, the NK model is a valuable framework with which to investigate numerous questions about the evolutionary dynamics of a system. Extensions to the core model, such as the model involving coupled landscapes, hold even deeper promise. *Origins of Order* may seem like a tough read at first, but I encourage you to take the effort to learn from it, and also from the application papers being written that use it in your specific domain of interest.

General references
Kauffman, S., (1993), *Origins of Order*, NY: Oxford University Press.
Kauffman, S., (1995), *At Home in the Universe*, NY: Oxford University Press.
Kauffman, S., (2000), *Investigations*, NY: Oxford University Press.

Economic and management applications
Choi, T., and K. Dooley (2000), "Conceptualizing Supply Networks As A Complex System: Its Meaning, Its Properties, And Managerial Implications, *J. of Operations Management* 19:351-366.
Kauffman, S. (1988), "The Evolution of Economic Webs", *The Economy as an Evolving Complex System*, Santa Fe Institute in the Sciences of Complexity, Ed. P.W. Anderson, K.J. Arrow and D. Pines, (Reading, MA: Addison-Wesley), vol. V, pp. 125.
Kauffman, S., Lobo, J., and W. Macready (1998), "Optimal Search on a Technology Landscape," Santa Fe Institute Working Paper 98-10-091.
Levinthal, D.A. (1997), "Adaptation on rugged landscapes", *Management Science* 43, p. 934.
Levinthal, D., and Warglien, M. (1999), "Landscape Design: Designing for Local Action in Complex Worlds," *Organization Science*, 10(3).
McKelvey, W. (1999), "Avoiding Complexity Catastrophe in Coevolutionary Pockets: Strategies for Rugged Landscapes," *Organization Science*, 10(3), 294-321.

Some web sites
http://cs.gmu.edu/~mpotter/nk-generator/
http://www.dcs.port.ac.uk/~hirsta/rugged.htm
http://www.ossu.co.uk/manstrategy/index.shtml
http://public.logica.com/~stepneys/bib/nf/kauffman.htm
http://www.marlboro.edu/~lmoss/planhome/thesis/litrev/02.evcomp.html
http://www.plexusinstitute.com/edgeware/archive/think/main_app-h.html
http://home.pscw.uva.nl/lleydesdorff/casys01/