Markov Systems and Symbolic Dynamics: Investigation 1

Bruce D. Marron
SYSC 505, Portland State University

June 13, 2015

Introduction

Science advances through perception. Direct perception is data collection coupled to processes of observation; indirect perception is causal analysis coupled to processes of experimentation, data analysis, and hypothesis testing. Taken together, direct and indirect perception can lead to real understanding of the universe we live in and is the reason why empirical science continues to add valuable knowledge to the human stock. It is noteworthy that the prowess of direct perception grows as a natural consequence of scientific investigation itself, which constantly creates new lenses with which to see the universe (new technologies), while the potency of indirect perception increases through the use of good experimental design and the optimal processing of incomplete information through inference [Jaynes and Bretthorst, 2003]. Thus scientific investigation can be considered as formalized perception, predicated on well-defined problem statements. As Haefner [Haefner, 2005] points out, problem statements are often translated from conceptual or mental scientific models to formal or mathematical scientific models. Such formal models may be considered as hypotheses and the tools of inferential statistics then may be used for hypothesis testing. In this way an understanding of the underlying mechanisms and processes that created the observed patterns can be abduced. Clearly, scientific problem statements assume the existence of some physical system to which they are causally linked. Science can thus be seen as a formalized heuristic for the perceptual investigation of user-defined systems where a system (after Zwick (2014), unpublished)

- is a set of elements having attributes linked by relations,
- is distinct and distinguishable from its environment,
- has internal sub-systems (organized parts) which constitute its structure, and
- participates in some external order (supra-system) which constitutes its function.

Science proceeds under the assumptions that (1) the system under investigation is nearly decomposable [Simon, 1996], and (2) the formal model is at least homomorphic to the system under investigation [Ashby, 1955]. Data are collected given these assumptions. Subsequent data analysis, as a critical component of causal analysis, typically refers to the use of both descriptive and inferential statistics. It is not inappropriate to state that descriptive statistics are, in fact, pattern recognition and pattern summary tools [Jain et al., 2000] while inferential statistics are tools for inductive reasoning [Devore, 2012]. Standard descriptive and inferential statistics are generally useful only for atemporal and non-spatial datasets. These are static datasets in the sense that they provide, at best, useful snapshots of underlying stochastic processes. Data analysis becomes much more difficult when the dimension of time (or space) is added and the data are time-based observations of real-world processes, both natural and artificial. That is, the data are now direct observations of the outputs of dynamic systems [Luenberger, 1979]. Such datasets are sequences (i.e., data points indexed over time) and time series statistical models such as autoregressive (AR) models, autoregressive moving average (ARMA) models, and state space models are often applied [Shumway, 2011]. Ultimately, data analysis is linked to causal, and not just logical inference because science seeks to understand the principles and mechanisms that actually produced the perceived data patterns.

Science, as sketched above, can be fully realized with remarkable results (in general) for conservative, machine-like systems in fields such as chemistry, hydrology, astrophysics, engineering, etc. where complexity in the system exists because of the evolution of matter and energy in time [Mainzer, 2007]. For dissipative systems like ecosystems, which evolve matter, energy, and information in time and across space, the complexity of the investigated system can be staggering [Mitchell, 2009]. Nevertheless, science has begun to build the perceptual and theoretical equipment needed to understand these complex adaptive systems.

This report documents the results from preliminary studies into the use of Markov chains for stochastic processes modeling. The intent of these studies is, ultimately, to develop a robust, applied methodology using symbolic dynamics. Such a methodology could, for example, be applied to the evaluation of the patterned heterogeneity in landscape ecology helping to sort out the pattern:process relationships that underscore landscape sustainability [Musacchio, 2009].

Background

Markov chains define discrete-state and discrete-step stochastic processes with an observable output that is an indexed sequence of uniquely defined states. The stochastic process generating such an observable sequence results from a system that has the ability to generate an observable, discrete random variable with at least two nominal values (i.e., two states). In the simplest case, the number of possible states is relatively small (fixed and finite) and the realization of the discrete random variable at some time, t, is dependent only on the realization of the random variable at the preceding time step. The dependency takes the form of state transition probabilities and so the process moves stepwise but randomly among the finite number of states. More formally, let

$$\begin{aligned}
\chi \equiv &\;\; an\ alphabet\ of\ states,\ s = 1, 2, 3, ..., n \\
|\chi| \equiv &\;\; the\ cardinality\ of\ \chi \\
X^{(s)} \equiv &\;\; a\ discrete\ random\ variable\ able\ to\ realize\ states,\ s \\
\{x_i^s\} \equiv &\;\; the\ sequenced\ realization\ of\ X^{(s)},\ i = 0, 1, 2, ..., t \\
[P_{s,s}] \equiv &\;\; a\ probability\ transition\ matrix \\
\equiv &\;\; Pr(s \to s)\ \ for\ all\ s\ in\ \chi \\
Pr(\{x_i^s\}) \equiv &\;\; the\ joint\ probability\ mass\ function\ of\ \{x_i^s\}
\end{aligned}$$

then a discrete stochastic process is said to be a Markov chain if for $|\chi| \geqslant 2$ and
i = 0,1, ..., t,

$$Pr(x_{i+1}^s|x_i^s, x_{i-1}^s, x_{i-2}^s, ..., x_0^s) = Pr(x_{i+1}^s|x_i^s)$$

and

$$Pr(\{x_i^s\}) = Pr(x_i^s|x_{i-1}^s)Pr(x_{i-1}^s|x_{i-2}^s), ..., Pr(x_0^s)$$

A Markov chain is said to be *stationary* (*time invariant*) if the conditional probabilities do not depend on t; that is, for i = 0,1,2, ..., t,

$$Pr(x_{i+1}^{s=b}|x_i^{s=a}) = Pr(x_1^{s=b}|x_0^{s=a}) \quad for\ all\ a, b\ in\ \chi$$

and a stationary, finite Markov chain is characterized by its initial state, $x_0^{s=a}$, and its probability transition matrix, $[P_{s,s}]$. Markov chains can successfully model chaotic dynamics if (1) the individual outcomes of a stochastic process can be defined deterministically as an indexed set of discrete, accessible states $\{S_i\}$ thereby partitioning the state space into discrete, exhaustive and mutually exclusive cells, and (2) the individual states are each uniquely identifiable [Nicolis and Prigogine, 1989]. The time evolution of the stochastic process thus produces sets of time series data,

$$S_1, S_2, ..., S_t$$

Remarkably, such a time series output is an asymmetric (irreversible), information-rich structure; a succession of "letters" of an "alphabet" that can be analyzed using the quantitative product rule of probability theory. Perhaps even more surprising is the fact that the asymptotic equipartition property of information theory applies and the set of all possible sequences generated can be divided into a typical set, where the sample entropy is close to the true entropy, and the nontypical set of all other sequences [Cover and Thomas, 2006].

General Approach

The development of a new methodology must start with plausible foundations. In the present case, the first step is to determine the plausibility of symbolic analysis on data generated from a known source. Put another way, this investigation seeks an answer to the question, "If a complex system is modeled as a simple Markov chain, is it possible to recover the underlying transition matrix from the empirical (time series) data?" The general approach taken for the current investigation is thus,

1. Define a complex dynamic system as a stochastic process generator.
2. Define the stochastic process as a two-state, stationary, irreducible Markov chain.
3. Use the transition matrix of the Markov chain as a time series data generator to simulate a sequence of symbols from a limited alphabet.
4. Analyze the time series data for the conditional probabilities of various alphabet combinations (symbolic pattern analysis).
5. Reconstruct the transition matrix from the data.
6. Evaluate the methodology by comparing its results to the known dynamic system.

Experimental

This section provides the methodological, procedural, and experimental details of the completed simulation work. Because full accountability is paramount for simulation studies and experiments (it assures that such work is transparent, reproducible, and not supported by uncheckable assumptions or assumptions that are not generally agreed upon), this section opens with the details of the computer hardware and software used. Note that this report includes a separate Appendix document where the details of all calculations, programming calls, models, and data processing are available. The Appendix includes all of the scripts necessary to reproduce any of the reported results (Experiment1b, Experiment2b, Experiment3b, Experiment4a).

Hardware, Base Operating System, and Software

All simulation experiments and subsequent data processing were performed on a Compaq 6710b laptop computer with the following hardware configuration:

```
Architecture:           i686
cpu op-mode(s):         32-bit, 64-bit
vendor_id:              GenuineIntel
cpu family:             6
```

```
version:               6.7.6
model:                 23
model name:            Intel(R) Core(TM)2 Duo CPU T8100 @ 2.10GHz
cpu(s):                2
cpu MHz:               2101.000
cache size:            3072 KB
```

The base operating system for the Compaq 6710b at the time the experiments were performed was,

```
Linux version:      3.2.0-85-generic, #122-Ubuntu SMP
Distributor ID:     Ubuntu
Description:        Ubuntu 12.04.5 LTS
Release:            12.04
Codename:           precise
```

All simulations and subsequent data processing and analyses in support of this project were performed in R, the open-source statistical program,

```
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: i686-pc-linux-gnu (32-bit)
```

Methods and Procedures

The methods and procedures used for this investigation are presented below in an algorithmic format as a sequence of steps. Each step is fully described with sufficient implementation details to provide an explanation of its purpose. The actual implementation details (scripts) and subsequent outputs (calculations) for each step are available in the Appendix.

Step 1: Definition of the dynamic systems

The investigation uses a two-state Markov chain as the basic generative model (Figure 1). Four generative models (G1,G2,G3,G4) were defined by their respective transition matrices as,

```
G1
     [,1] [,2]
[1,]  0.9  0.1
[2,]  1.0  0.0
```

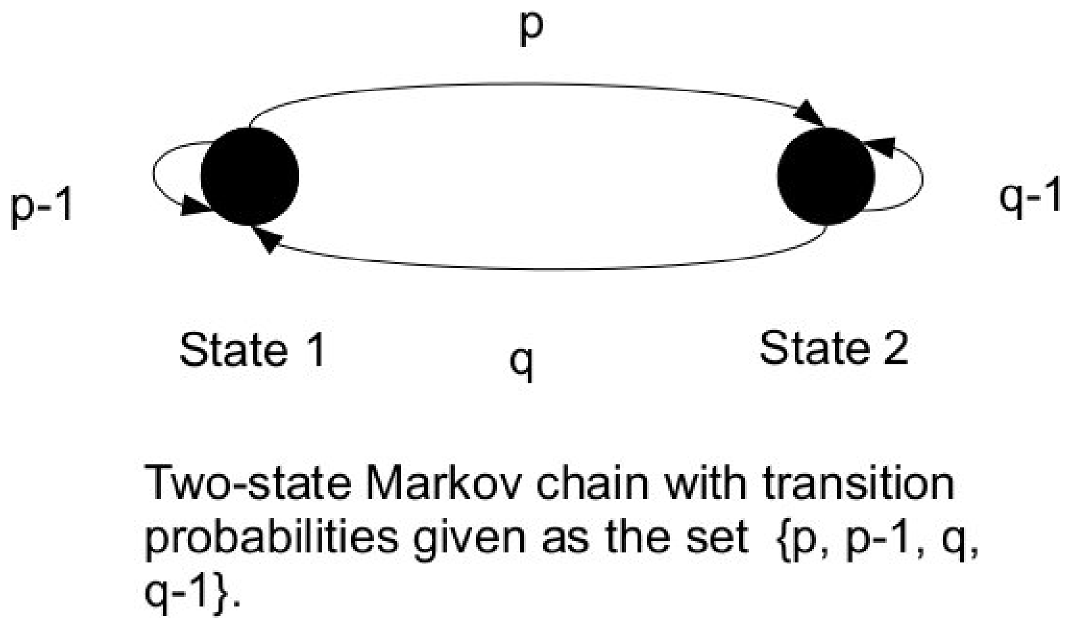
```
G2
     [,1] [,2]
[1,]  0.9  0.1
[2,]  0.5  0.5

G3
     [,1] [,2]
[1,]  0.5  0.5
[2,]  0.5  0.5

G4
     [,1] [,2]
[1,]  0.9  0.1
[2,]  1.0  0.0
```

where

$$\begin{aligned}
G_i[1,1] &\equiv a\ transition\ from\ State\ 1\ to\ State\ 1 \\
G_i[1,2] &\equiv a\ transition\ from\ State\ 1\ to\ State\ 2 \\
G_i[2,1] &\equiv a\ transition\ from\ State\ 2\ to\ State\ 1 \\
G_i[2,2] &\equiv a\ transition\ from\ State\ 2\ to\ State\ 2
\end{aligned}$$

All of the generative models (Markov chains) are regular and irreducible, and all were determined to converge to a stationary distribution using the method of consecutive matrix powers [Luenberger, 1979]. The method of consecutive matrix powers raises the transition matrix to ever-increasing powers until two consecutive results match a user-defined criterion. Here, a transition matrix was considered stationary if two consecutive powers gave the exact same values rounded to five decimal places.

Figure 1: The generative model.



Two-state Markov chain with transition probabilities given as the set {p, p-1, q, q-1}.

Step 2: Generation of the simulated datasets

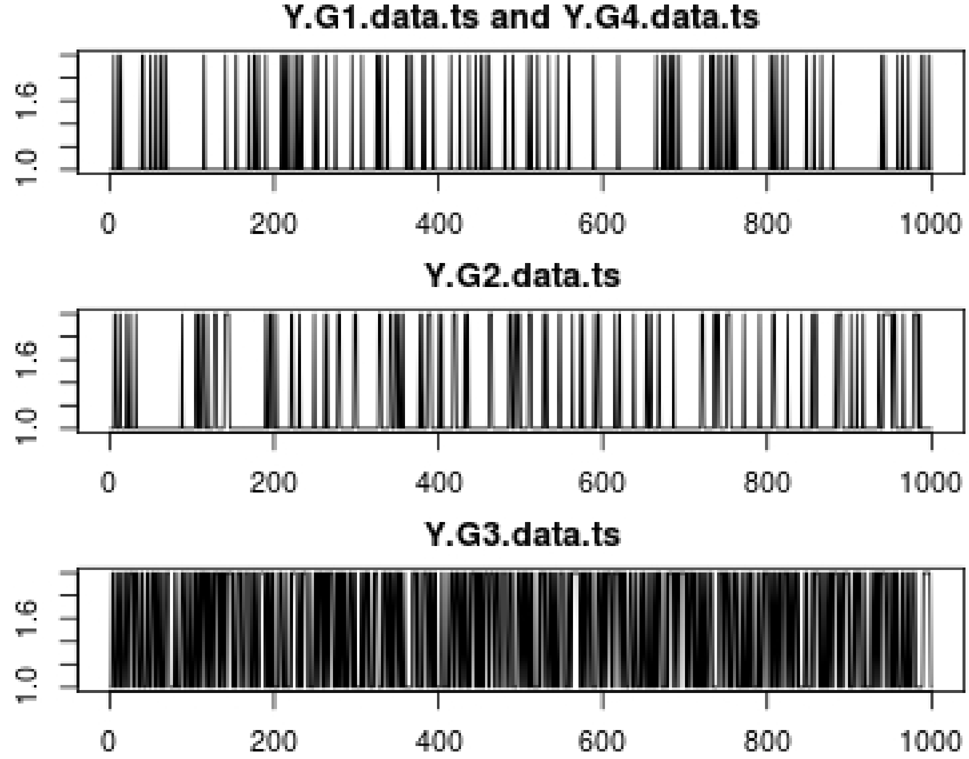
A simple variant of the Metropolis-Hastings algorithm was created and then implemented to produce Monte Carlo datasets (n=1000) using a Gibbs-like sampler[Hoff, 2009][MacKay, 2003]. Basically, the sampler draws random deviates from the transition matrix conditional on the current state of the system. Such sampling directly simulates the observational output from the generative model and produces simulated time series data (sequences). The sampler for generative model G1 along with the first 50 data points from each generative model are given as,

```
---- sampler for generative model G1 --------------
library(distr)
exp1gen1d1 <- DiscreteDistribution (supp = c(1, 2) , prob = c(0.9, 0.1))
exp1gen1d2 <- DiscreteDistribution (supp = c(1, 2) , prob = c(1, 0))
set.seed(74)
tiks<-1000
Y_0<-1                          #at t=0 ==> State 1
Y.G1.data <- NULL               #empty vector
Y.G1.data[1] <- Y_0
```

```
for (i in 2:tiks) {

if (Y.G1.data[i-1]=="1"){
 Y.G1.data[i] <- r(exp1gen1d1)(1)
} else if (Y.G1.data[i-1]=="2"){
Y.G1.data[i] <- r(exp1gen1d2)(1)
}
}

---- simulated time series data from the sampler ----------------
  Y.G1.data and Y.G4.data
   [1] 1 1 1 1 2 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
  [39] 2 1 2 1 1 1 1 1 1 1 2 1 ...

  Y.G2.data
   [1] 1 1 1 1 1 2 2 1 1 1 1 2 2 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1
  [39] 1 1 1 1 1 1 1 1 1 1 1 1 ...

  Y.G3.data
   [1] 1 1 2 2 1 1 1 2 1 1 2 2 1 1 2 2 1 1 1 2 2 1 1 2 2 2 2 1 2 1 2 2 2 1 2 1 1 1
   [39] 2 2 1 1 1 1 2 2 1 2 1 1 ...
```

An interesting display of the data from the generative models is given in Figure 2 below.

Figure 2: Line plots of the simulated datasets produced by generative models G1, G2, G3, and G4.



Step 3: Determination of the conditional probabilities and transition matrices from the probabilities of subsequences in the simulated datasets

Symbolic dynamics analysis is a type of pattern analysis that, similar to latent variable techniques, assumes (1) that there is a deterministic but unknown process operating that is responsible for generating the sequences, and (2) that strong correlation is to be expected in the succession of symbols generated by a Markov process [Nicolis and Prigogine, 1989]. The subsequences of interest are the combinations of letters of the alphabet (states) that are needed to derive conditional probabilities from the basic product rule of probability theory. So for example, the conditional probability of State 1 given State 2 can be coded as,

$$Pr(State\ 1|State\ 2) \equiv\ p.a\_b$$

where,

$$p.a\_b \quad = \frac{p.ba}{p.b}$$

If the counts of the subsequences {b} and {ba} can be determined from the data, then values for *p.b* (= the frequency of {b} in the data) and *p.ba* (= the frequency of {ba} in the data) can be calculated. Subsequently the conditional probabilities and the empirical transition matrices can be determined. Finding the counts of singlets ({a, b}) in the data is straightforward; finding the counts for doublets ({aa, ab, ba, bb}) and triplets ({aaa, aab, aba, abb, baa, bab, bba, bbb}) required the creation of moving window functions. For example, the script used to count doublets in the data is given as,

```
count <- NULL
for (i in 1:999) {
            #doublet {11} ==> "1"
if(identical(window(Y.G1.data.ts, i, i+1)[1:2], c(1,1))){
count[i]<-1
}else{
           #doublet {12} ==> "2"
if(identical(window(Y.G1.data.ts, i, i+1)[1:2], c(1,2))){
count[i]<-2
}else{
           #doublet {21} ==> "3"
if(identical(window(Y.G1.data.ts, i, i+1)[1:2], c(2,1))){
count[i]<-3
}else{
           #doublet {22} ==> "4"
if(identical(window(Y.G1.data.ts, i, i+1)[1:2], c(2,2))){
count[i]<-4
}
      }
    }
  }
}
```

Step 4: Determine the empirical transition matrices and their stationary distributions

Determination of empirically-derived transition matrices and their stationary distributions through the use of symbolic dynamics is the primary goal of this investigation. Although there may be more accurate ways of determining the empirical transition matrix, a legitimate 'first-cut' is simply to use the appropriate conditional probabilities directly. Thus,

$$\begin{aligned}
empirical\ G_i[1,1] &\equiv p.a\_a \\
empirical\ G_i[1,2] &\equiv p.b\_a \\
empirical\ G_i[2,1] &\equiv p.a\_b \\
empirical\ G_i[2,2] &\equiv p.b\_b
\end{aligned}$$

Determination of the stationary distributions was performed by the method iterative matrix powers as described above.

Step 5: Evaluation of methodology performance

The evaluation of basic methodology performance includes (1) direct comparison of the methodology-derived transition matrices to the known transition matrices, (2) comparison of the methodology-derived stationary distributions to the known stationary distributions using the Kullback-Leibler divergence (relative entropy) [MacKay, 2003], (3) evaluation of the methodology-derived joint probability distributions using the asymptotic equipartition theorem for typical sets [Cover and Thomas, 2006], and (4) exploration of the methodology through the use of a bootstrap sampling distribution (Experiment4a)[Lunneborg, 2000].

Results and Discussion

The evaluation of the methodology-derived transition matrices is summarized in Table 1; the evaluation of the methodology-derived stationary distributions is summarized in Table 2. Table 3 and Table 4 summarize the dissimilarity between the probability of a sequence given its joint probability distribution and the probability of a typical sequence as given by the asymptotic equipartition theorem. Figure 3, Figure 4, and Figure 5 display the time lag plots of the sample autocorrelation function for generative models G1, G2, and G3, respectively [Shumway, 2011].

There is much to digest here and, regrettably a complete and detailed analysis of the methodology must wait[1]. Nonetheless, it is clear that the methodology holds promise: it seems possible to use symbolic dynamics to derive a good estimate of the transition matrix for stochastic process governed by a two-state Markov chain, even if bootstrapping is required because of a small sample size. Should the methodology pan out after detailed analysis and thorough experimentation, I foresee huge utility in its application. For example, if there is a way to define a basic discrete state space, the method can be extended with the use of arithmetic codes to handle complex adaptive models [MacKay, 2003]. In fact, it may even be possible to develop a synoptical key for complex systems analysis using this methodology.

[1] Pending engagements and current deadlines demand that I give only a cursory analysis at this time.

Figure 3: Autocorrelation of simulated time series data from generative model G1.

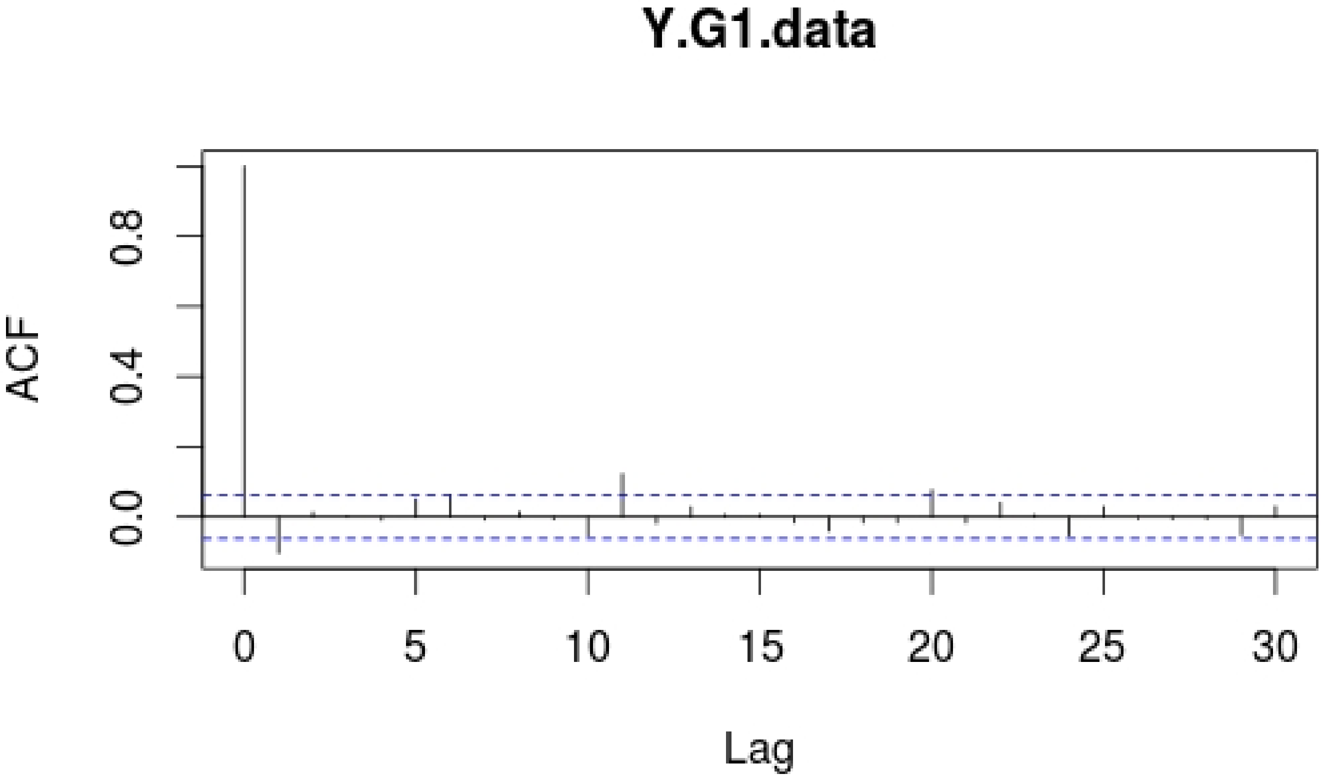


Table 1: The known transition matrices of the simulated (Monte Carlo) data, the transition matrices as derived by the method of symbolic analysis, and the mean parameter accuracies as mean percent relative error.

Generator	Known Trans. Matrix	Derived Trans. Matrix	Mean Percent Rel. Error
G1	[0.9, 0.1, 1.0, 0.0]	[0.8960, 0.1040, 1.0000, 0.0000]	1.19
G2	[0.9, 0.1, 0.5, 0.5]	[0.9040, 0.0960, 0.4730, 0.5270]	-0.89
G3	[0.5, 0.5, 0.5, 0.5]	[0.5115, 0.4885, 0.4965, 0.5035]	0.00
G4	[0.9, 0.1, 1.0, 0.0]	[0.8994, 0.1006, 1.0000, 0.0000]	0.18

Figure 4: Autocorrelation of simulated time series data from generative model G2.

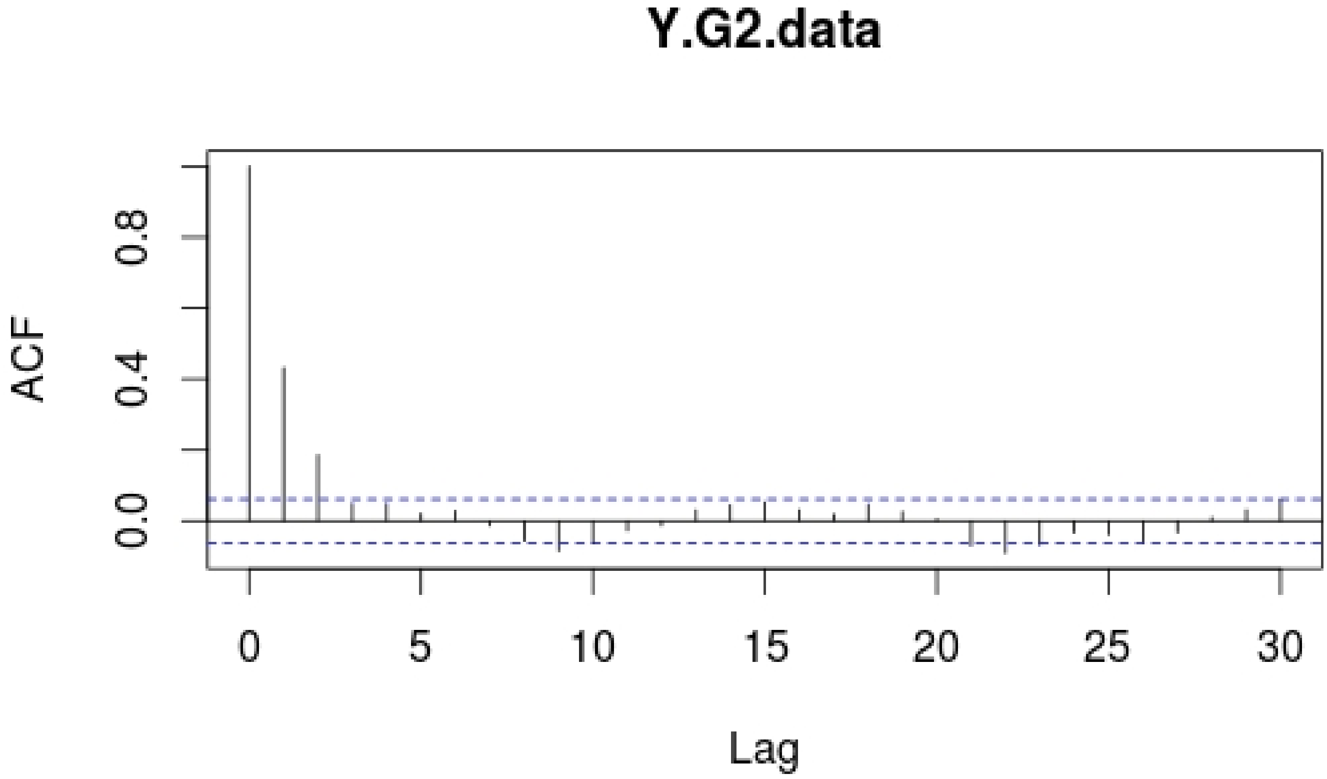


Table 2: The known stationary distributions of the simulated (Monte Carlo) data, the stationary distributions as derived by the method of symbolic analysis, and their relative entropy (Kullback-Leibler divergence).

Generator	Known Stat. Dist.	Derived Stat. Dist.	Relative Entropy
G1	[0.909, 0.091]	[0.906, 0.094]	7.70e-05
G2	[0.833, 0.167]	[0.831, 0.169]	2.06e-05
G3	[0.500, 0.500]	[0.504, 0.496]	4.62e-05
G4	[0.909, 0.091]	[0.910, 0.090]	8.78e-06

Figure 5: Autocorrelation of simulated time series data from generative model G3.

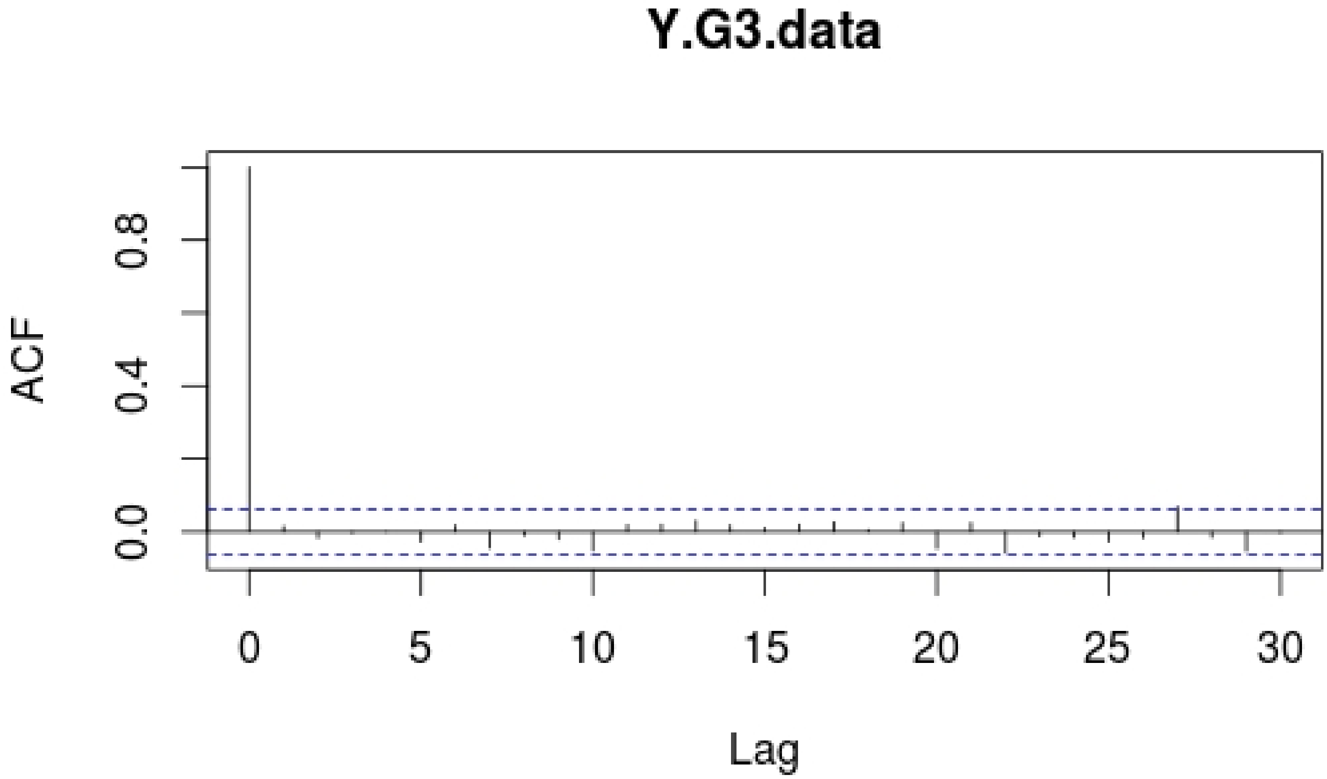


Table 3: The probability of simulated sequences given their known joint probability mass function, the asymptotic equipartition (AE) value for the probability of simulated sequences in the typical set given the known stationary distribution, and a dissimilarity measure (the absolute value of the log of their ratio).

Generator	Known Seq. Prob.	Known AE Seq. Prob.	Dissimilarity
G1	7.77e-132	4.05e-133	4.26
G2	1.87e-164	1.23e-196	106.
G3	1.87e-301	9.33e-302	1.00
G4	1.33e-06	2.40e-07	2.47

Table 4: The probability of simulated sequences given their methodology-derived joint probability mass function, the asymptotic equipartition (AE) value for the probability of simulated sequences in the typical set given the methodology-derived stationary distribution, and and a dissimilarity measure (the absolute value of the log of their ratio).

Generator	Derived Seq. Prob.	Derived AE Seq. Prob.	Dissimilarity
G1	8.37e-132	4.287e-136	14.3
G2	2.67e-164	5.03e-198	112.
G3	2.17e-301	9.64e-302	1.17
G4	1.33e-06	2.69e-07	2.31

Conclusion

Generative models, as compared to statistical models, seem much more likely to help uncover real causal relationships in systems under scientific investigation because they can, ideally, simulate the system's observable output (behavior). Plausible reasoning (inference) from such models seems a bit less ad hoc than is often the case for Fisherian statistical inference. If this methodology could be applied within a synoptical key framework, it would help to resolve the choice of scale and system:subsystem definition issues that plague attempts to study complex adaptive systems. Such issues are, for example, at the forefront of sustainable landscape ecology.

References

[Ashby, 1955] Ashby, W. R. (1955). *An introduction to cybernetics.* Taylor & Francis.

[Cover and Thomas, 2006] Cover, T. M. and Thomas, J. A. (2006). *Elements of information theory.* Wiley-Interscience, Hoboken, N.J, 2nd ed edition.

[Devore, 2012] Devore, J. L. (2012). *Probability and statistics for engineering and the sciences.* Brooks/Cole, Cengage Learning, Boston, MA, eighth edition edition.

[Haefner, 2005] Haefner, J. W. (2005). *Modeling biological systems: principles and applications.* Springer, New York, 2nd ed edition.

[Hoff, 2009] Hoff, P. D. (2009). *A first course in Bayesian statistical methods.* Springer texts in statistics. Springer, London ; New York.

[Jain et al., 2000] Jain, A. K., Duin, R. P. W., and Mao, J. (2000). Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1):4–37.

[Jaynes and Bretthorst, 2003] Jaynes, E. T. and Bretthorst, G. L. (2003). *Probability theory: the logic of science.* Cambridge University Press, Cambridge, UK ; New York, NY.

[Luenberger, 1979] Luenberger, D. G. (1979). *Introduction to dynamic systems: theory, models, and applications.* Wiley, New York.

[Lunneborg, 2000] Lunneborg, C. E. (2000). *Data analysis by resampling: concepts and applications.* Duxbury, Australia ; Pacific Grove, CA.

[MacKay, 2003] MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms.* Cambridge University Press, Cambridge, UK ; New York.

[Mainzer, 2007] Mainzer, K. (2007). *Thinking in complexity: the computional dynamics of matter, mind, and mankind.* Springer complexity. Springer, Berlin ; New York, 5th rev. and enl. ed edition.

[Mitchell, 2009] Mitchell, M. (2009). *Complexity: a guided tour.* Oxford University Press, Oxford [England] ; New York.

[Musacchio, 2009] Musacchio, L. R. (2009). The scientific basis for the design of landscape sustainability: a conceptual framework for translational landscape research and practice of designed landscapes and the six Es of landscape sustainability. *Landscape Ecology*, 24(8):993–1013.

[Nicolis and Prigogine, 1989] Nicolis, G. and Prigogine, I. (1989). *Exploring complexity : an introduction.* W.H. Freeman, New York.

[Shumway, 2011] Shumway, R. H. (2011). *Time series analysis and its applications: with R examples.* Springer texts in statistics. Springer, New York, 3rd ed edition.

[Simon, 1996] Simon, H. A. (1996). *The sciences of the artificial.* MIT Press, Cambridge, Mass, 3rd ed edition.

Appendix

Due to size, the Appendix for this report is contained in a separate document. The Appendix contains the following files: (1) Experiment1b.pdf, (2) Experiment2b.pdf, (3) Experiment3b.pdf, (4) Experiment4a.pdf, and (5) Evaluation1.pdf.