

Quantitative Research Methods in Chaos and Complexity: From
Probability to Post Hoc Regression Analyses

Donald L. Gilstrap – Wichita State University

Deposited 11/15/2018

Citation of published version:

Gilstrap, Donald L. (2013): Quantitative Research Methods in Chaos and Complexity: From Probability to Post Hoc Regression Analyses. *Complicity: An International Journal of Complexity and Education*, 10 (1/2).

<https://journals.library.ualberta.ca/complicity/index.php/complicity/article/view/20400>

© Copyright 2013. The author, DONALD GILSTRAP, assigns to the University of Alberta and other educational and non-profit institutions a non-exclusive license to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The author also grant a non-exclusive license to the University of Alberta to publish this document in full on the World Wide Web, and for the document to be published on mirrors on the World Wide Web. Any other usage is prohibited without the express permission of the authors.

Quantitative Research Methods in Chaos and Complexity: From Probability to Post Hoc Regression Analyses

DONALD L. GILSTRAP
Wichita State University, (USA)

In addition to qualitative methods presented in chaos and complexity theories in educational research, this article addresses quantitative methods that may show potential for future research studies. Although much in the social and behavioral sciences literature has focused on computer simulations, this article explores current chaos and complexity methods that have the potential to bridge the divide between qualitative and quantitative, as well as theoretical and applied, human research studies. These methods include multiple linear regression, nonlinear regression, stochastics, Monte Carlo methods, Markov Chains, and Lyapunov exponents. A postulate for post hoc regression analysis is then presented as an example of an emergent, recursive, and iterative quantitative method when dealing with interaction effects and collinearity among variables. This postulate also highlights the power of both qualitative and quantitative chaos and complexity theories in order to observe and describe both the micro and macro levels of systemic emergence.

Introduction

Chaos and complexity theories are numerous, and, for the common reader of complexity texts, it is easy to develop a personal panacea for how chaos and complexity theories work. However, this journal highlights the different definitions and approaches educational researchers introduce to describe the phenomena that emerge during the course of their research. Converse to statements from outside observers, complexity theory in its general forms is not complicated, it is complex, and when looking at the micro levels of phenomena that emerge we see much more sophistication or even messiness. It draws from numerous observations and analyses, constructing a more holistic interpretation of social events regarding teaching and learning. Certainly debate has ensued among this group of scholars who write for *Complicity* for many years on whether method even has a place in chaos and complexity theories firmly rooted in the social and behavioral sciences. This special issue will draw out that continued dialogue.

In qualitative research in the social sciences, a growing body of research methods is developing that can be integrated into educational research frameworks (Hesse-Biber & Leavy, 2008). In quantitative research in the social sciences, little exists that posits

complexity theory in the framework of “numbers.” Although we often base our metaphorical lenses upon the theoretical constructs of physics, mathematics, and biology, we equally encounter the difficulties of reconciling methods we borrow from these disciplines that perhaps can form bridges to the educational research community in the study of humans. Amidst the numerous frameworks for understanding chaos and complexity theories, it is the purpose of this paper to focus specifically on quantitative methods. This article investigates the power of *post hoc* quantitative methods which are concurrently emergent, recursive, and iterative at the point of analysis, including probability analysis, regression, and other tools of analysis. Although space is limited for detailed examples of each method, I integrate a working example of *post hoc* analysis using a regression model. Moreover, this article suggests ways of viewing complexity methods in educational research as multi-faceted in providing a holistic view of chaos and complexity theories.

Quantitative methods

Multiple linear regression

Regression analysis has become an extremely popular tool for analysis in research methods involving multiple independent variables. The *regression* line, for which it is named, is composed of data points in two dimensional with one axis representing expected cumulative probabilities and the other axis representing observed cumulative probabilities. Multiple linear regression is not a good fit with chaos and complexity theories, because it follows the regression line in a linear format in order to arrive at generalization (figure 1). Oftentimes it is used as a model for prediction where Analysis of Variance (ANOVA) or Analysis of Co-Variance (ANCOVA) would serve this purpose in much better form, since ANOVA and ANCOVA are designed specifically for tests of significance of variables as predictors of dependent variables. However, multiple regression has often been emphasized by statisticians in the social and behavioral sciences as a model that should be used to confirm or explain phenomena or even to build theory. Therefore, intrinsically, regression analysis at the surface provides great potential for chaos and complexity theories research methods, because the model incorporates a large number of variables, can handle different types of variables from categorical to interval, calculates interaction effects between variables, and, when used appropriately, is confirmatory or explanatory in nature. Extending the concept of interaction effects, moderated multiple regression takes into account one or more moderator variables that change the influences of the independent variables on dependent variables, resulting in several different outcomes (Aguinis, 2004). A moderator variable, by definition, “affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable” (Baron & Kenny, 1986). In a nutshell, the moderator variable, serves as a bridge between an independent variable and a dependent variable. As an example, in the case of a person’s political party preference influencing attitudes toward children’s social programs, gender may serve as a moderator variable, where women might support these programs regardless of their political affiliation. The very concept of moderation during multiple regression shows how complex interactions can be among and between different variables and, perhaps, reinforces how our knowledge of every variable *a priori* in human studies is extremely difficult if not impossible. However, regression analysis has to be used in different ways than we typically find in the research literature in order to incorporate the broad, rich, and diverse data sets that exist within the chaotic world.

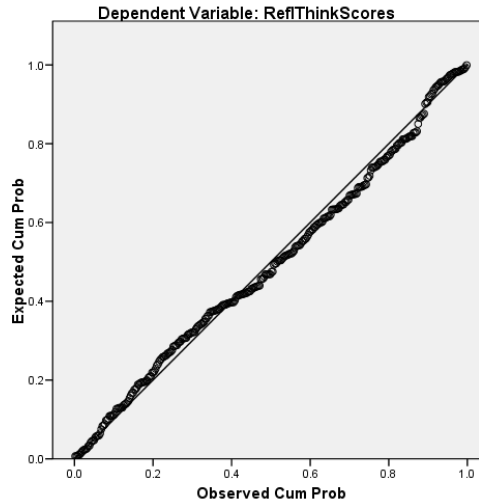


Figure 1: Regression Line

Nonlinear regression

Nonlinear regression perhaps moves us closer toward a tool for investigating complex and chaotic phenomena, since it is designed to explore curvilinear relationships. Whereas linear regression models of significance follow the regression line tightly, nonlinear regression models do not. They have a curvilinear relationship to the regression line and subsequently, to each variable in the model (figure 2). Rather than being specific to the point of reduction, nonlinear regression iteratively compares relationships among variables and computes correlations rather than predictions respectively. Although some researchers in the social and behavioral sciences might argue for data transformations or “quantifying the roughness of a curve” (Green & Silverman, 2000, p. 4) when data becomes messy, it is actually the nonlinear and nongeneralizable aspects of data in its raw forms that provide insights into quantitative applications in human studies. In time series analyses, Guastello (2011) notes that nonlinear regression is particularly useful in helping to discover which variable(s) contribute dramatically to bifurcation points through the incorporation of the Lyapunov exponent discussed later in this paper. Interestingly, Csikszentmihalyi’s (1990) work on *flow* was explored by Guastello, Johnson, and Rieke (1999), and these authors found that, by using nonlinear regression, one could show “that the level of flow experience is a chaotic process over long-enough periods of time” (Guastello, 2011, p. 289).

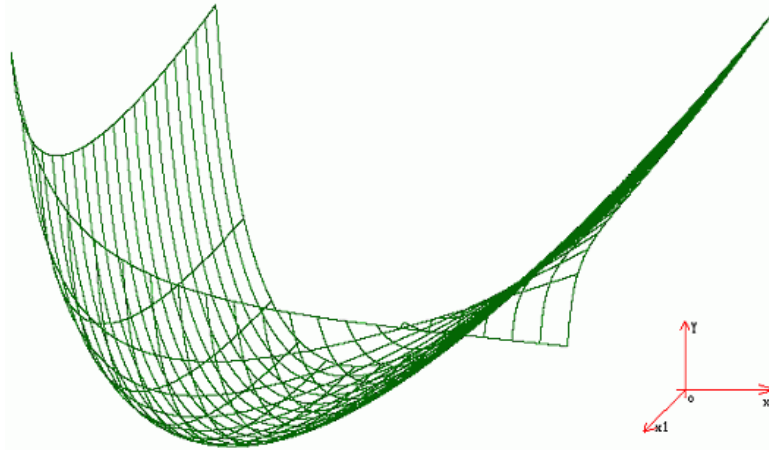


Figure 2: Nonlinear regression

Probability analysis

Stochastics

Chaotic systems differ from deterministic, even dynamical systems, in the amount of information that is required to observe and define chaos. As Stark (2001) observes, however:

Unfortunately, in many real problems one has only the sketchiest information about what these variables are, and one certainly has no hope of observing them all. Instead, one typically has a time series of one or more observables of the system, whose relationship to the state variables is at best uncertain. (Stark, 2001, p. 81).

And it is this relationship between known and unknown variables that lend stochastic models to research methods in chaos and complexity theories. Stochastics and probability analysis have been at the heart of research in chaos and complexity theories for many years. Gleick (1987) used the term to act as a bridge between determinism and chaos; Prigogine and Stengers (1984, p. 276; Prigogine, 1996) incorporated stochastics to characterize and distinguish “intrinsic randomness” and “intrinsic reversibility” from deterministic systems; Waldrop (1992) used observation in non-normal stochastic time series in economics to reveal complexity theory.

The power of stochastics methods for understanding chaotic and complex phenomena is that they can incorporate many variables in a phase space similar to what Prigogine (1996) describes, where the *arrow of time* is a contributing factor in understanding emergent phenomena. In a three dimensional model with the absence of time, phase spaces play no role. However, when the *arrow of time* is introduced into a four dimensional model, multiple phase spaces are present, and each phase space represents a possible trajectory of systemic outcomes. One might say that a stochastic model casts a net, bringing in many random variables while analyzing the influence of each variable, each subset of variables, and each group of variables between and among each other. A picture describing likely outcomes is then presented at the point of analysis (figure 3), where I would contend the deterministic chaos of Poincaré (1890) is then revealed. In his approach, certain boundary controlling conditions limit a chaotic system from ever reaching unbounded chaos, thereby showing that these particular system outcomes can be determined *a priori* to observations over time. My argument here is that deterministic chaos still avoids the richness of chaos theory, as it protects

potential outcomes from occurring by placing each variable within boundary conditions that rule out anomalies. Equally, when several stochastic models are applied to the same data sets, yet with different boundary parameters, then probability analysis becomes much richer by viewing outcomes at a macro level of emergent subsets. As an example, in figure 3 each elevation or declination in the model represents an emergent subset; however, a true picture of the systemic outcomes can only be seen when viewing all of the subsets simultaneously.

The downfall to stochastic methods, however, is that they need a starting point. In human studies, we have no choice as researchers but to subjectively choose that point as an encapsulation of one or more individuals' lives. Granted, this is often the case in much of the research we conduct in educational settings, whether qualitative or quantitative, but, at the same time, this paradox seems to be a critical element in understanding how chaotic and complex phenomena emerge. Equally, stochastics are not used to predict but are rather observations of momentum or critical moments in data sets to which we should pay attention.

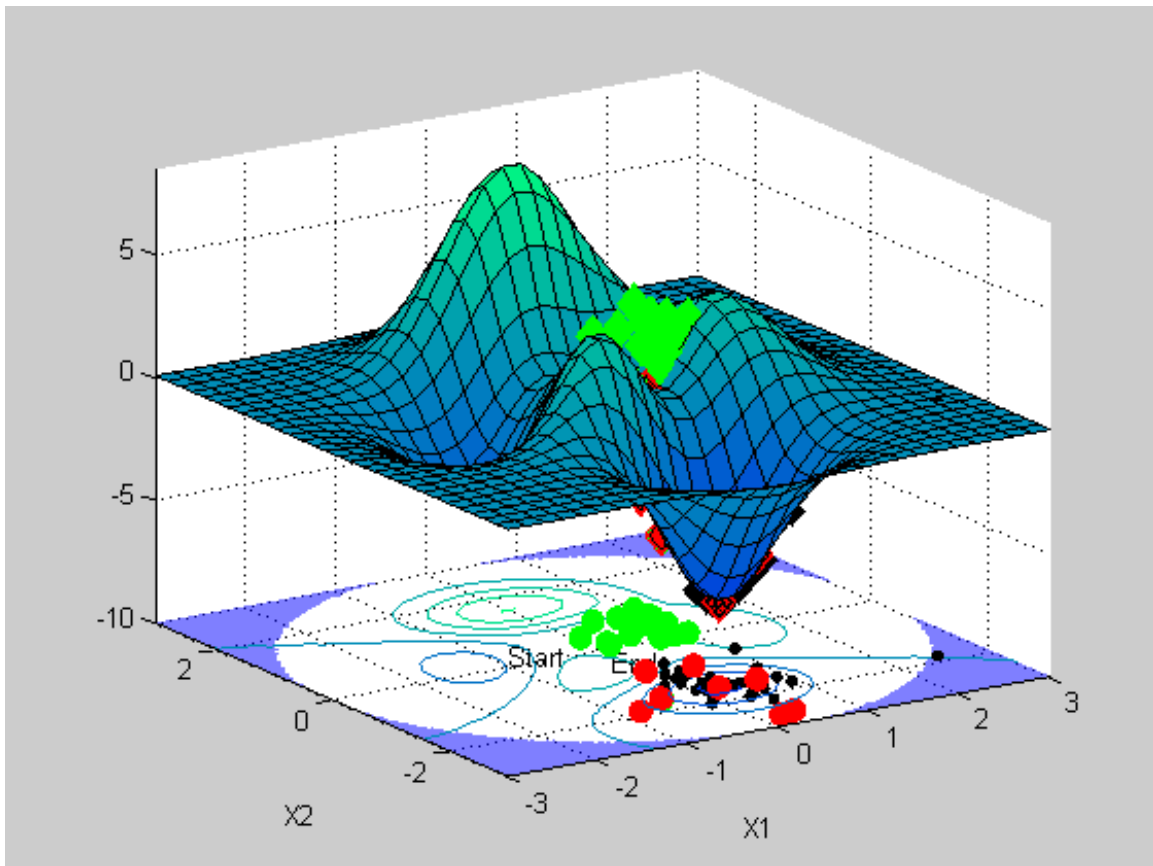


Figure 3: Stochastics

Monte Carlo methods

One commonly used quantitative research method, particularly for modeling and simulation, is the Monte Carlo method. Interestingly, the name Monte Carlo originated during the Los Alamos Laboratory experiments of the 1940s, where information was shrouded in secrecy. One of the researchers, Stanislaw Ulam, supposedly chose this pseudonym, since his uncle use to like to gamble in the casinos of Monte Carlo (Grinstead & Snell, 1997; Metropolis & Ulam, 1949). The difficulty of linear probability

analysis is that it usually takes into account random variables on one linear trajectory. The Monte Carlo Method utilizes several sets of random variables from many different trajectories over a period of time and then calculates probability outcomes. One of the benefits of this method is that it allows the researcher to average more accurately and more efficiently different system states over time. In computer science, this type of modeling is particularly useful, since, during the Gaussian [normal] distribution, large and frequently dispersed integrands (e.g. computer programming code functions that need to be integrated in order to create code libraries), are difficult to compute. Monte Carlo provides a way to average data at probabilistic intervals over time rather than through discrete time units. Equally, in cases where a large number of variables are difficult to include in mainstream quantitative models, Monte Carlo provides a means to randomly sample from or randomly generate variables of probable interest over time (Bar-Yam, 1997; Lemieux, 2009; Liang, Liu, & Carroll, 2010).

Since Monte Carlo methods seem particularly well suited for large data sets and computer simulated modeling – which we have seen overwhelmingly in the largest body of research on quantitative methods in chaotic and complex systems – I do not want to develop this in too much in detail. However, it is important to note that Monte Carlo methods provide a bridge between stochastics and Markov Chains, which will be discussed in the following section. We are able to integrate these methods into large-scale projects in education, from looking globally at education financing (Lemieux, 2009) to understanding large numbers of frequency distributions in an iterative fashion (Liang, Liu, & Carroll, 2010). Mortgage backed securities, which take 20 to 30 years to mature, are difficult to project with loglinear models because this long time span opens the model up to multiple emergent anomalies. Liemieux (2009) showed that the imminent failures of mortgage backed securities could also be seen, using more complex Monte Carlo analyses that would incorporate several different random variable sets at various states of a security's development, and subsequently making analysis much more robust and comprehensive. Likewise, frequency distributions are static in nature in linear models. However, Liang, Liu, and Carroll (2010) showed that, by using several samples from within a larger population sample, one can continually see emerging frequency distributions when they are iteratively drawn from a changing population sample over time.

Extending Monte Carlo methods to living systems, Bar-Yam (1997) poignantly addressed the application of the model when integrating evolutionary theory and phenomenology. In his view, we are able to observe phenomena and describe those systemic events through the concept of the fitness landscape. Those variables that have a high probability of influencing the evolution of a system can be randomly sampled over time, and subsequent correlations emerge. That being the case, we might distinguish between classical and quantum Monte Carlo methods. Classical methods seek out probability distributions. In some ways, we might be able to identify these same types of highly probable, influential variables in education – such as No Child Left Behind or consecutive years of difficult state budgets – and develop probability scenarios for coming years in how our educational systems may evolve. Quantum methods rely on a random walk through data over time, however. In research settings, we can randomly choose variables within large data sets, such as the National Center for Educational Statistics (NCES), and analyze historiographical phenomena that emerge over time (figure 4).

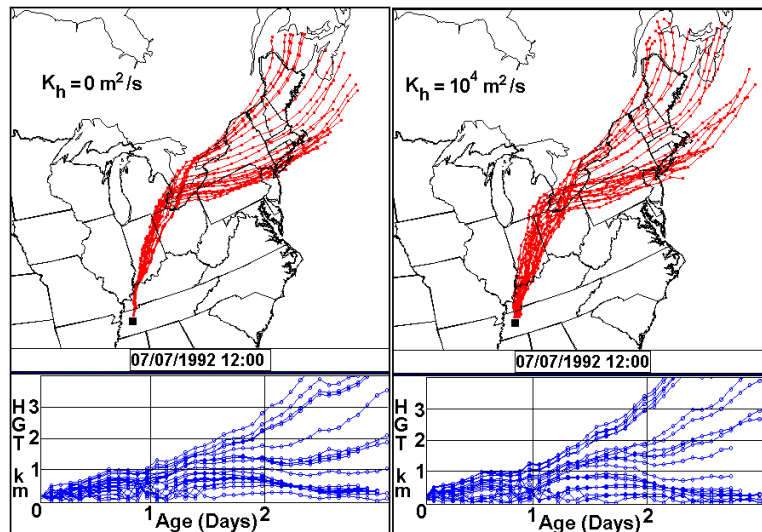


Figure 4: Monte Carlo Simulation of Air Pollution

Markov Chains

A model that is suitable for modelling a sequence of random variables... in which the probability that a variable assumes any specific value depends only on the value of a specified number of most recent variables that precede it. (NRG, 2004).

Russian mathematician Andrei Markov first applied the concept of Markov Chains during the early twentieth century in an analysis of Pushkin's *Eugene Onegin*. In a challenge to his opponent in theoretical mathematics, Pavel Nekrasov, Markov set out to prove that the laws of very large numbers were not subject to complete randomness (known at the time as the principle of independence) but rather they contained sets of probabilities. Using only pen and paper:

Markov's sample comprised the first 20,000 letters of the poem, which is about an eighth of the total... In the first phase of his analysis he arranged the text in 200 blocks of 10×10 characters, then counted the vowels in each row and column. From this tabulation he was able to calculate both the mean number of vowels per 100-character block and the variance, a measure of how widely samples depart from the mean... In a second phase Markov returned to the unbroken sequence of 20,000 letters, combing through it to classify pairs of successive letters according to their pattern of vowels and consonants. He counted 1,104 vowel-vowel pairs and was able to deduce that there were 3,827 double consonants; the remaining 15,069 pairs must consist of a vowel and a consonant in one order or the other. With these numbers in hand, Markov could estimate to what extent Pushkin's text violates the principle of independence. (Hayes, 2013, p. 95).

The legacy of this experiment continues, as this same algorithm is now used in computer modeling to determine both authorship and plagiarism of textual works.

Markov Chains are applied in contemporary social sciences research in numerous ways, ranging from Monte Carlo methods to Hidden Markov Models analyses. They are used to analyze a series of variables or events over time or during a discrete unit of time. Much like fractal iterations, Markov Chains also rely on the values of the previous system state to describe the new system state (figure 5). Moreover, Markov Chains take into account model noises that appear as residual outliers in the normal Gaussian distribution, where other models simply exclude that data as anomalies (Gregson, 2010). Additionally, this residual noise is nonlinear and can appear in the form of a basin of attraction, reinforcing the data's importance to remain within the model rather than removing it from the model. Equally important, Merrill (2010) proposes that Markov

Chains exhibit phenomena of bifurcations as observational representations of transitional changes from one system state to another. In psychology, Markov chain methods have been used to analyze mental states that fluctuate, such as bipolar disorders. In educational settings, we might also see application for this type of method to analyze learning and teaching phenomena such as critical reflection, where the learner's mental models fluctuate between a state of equilibrium and non-equilibrium (Gilstrap, 2010). In particular, because Markov methods include time series analysis, the implications for *post hoc* analyses can be instrumental in describing phenomena that appear in chaotic and iterative forms in educational settings.

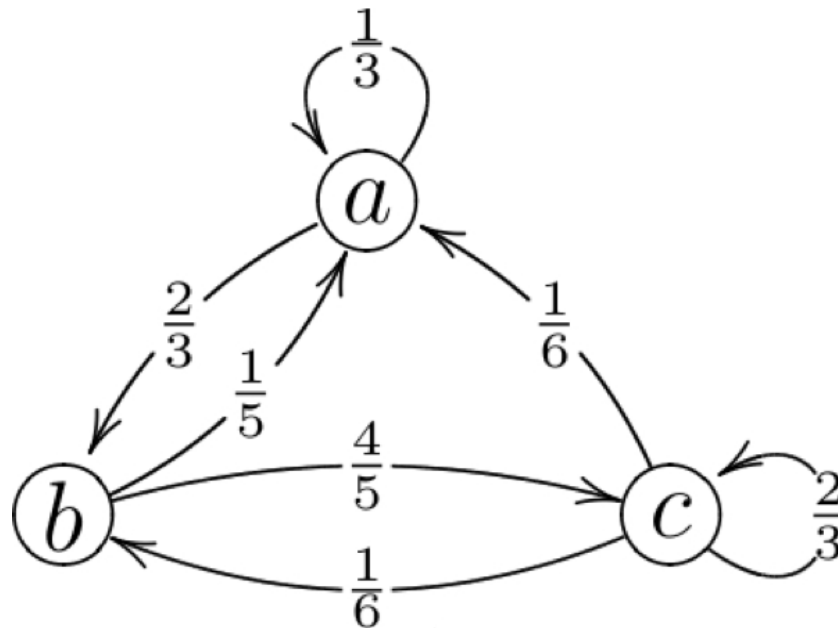


Figure 5: Markov Chain

Lyapunov exponent

If all points in a neighborhood of a trajectory converge toward the same trajectory, the attractor is a fixed point or a limit cycle. However, if the attractor is strange, any two trajectories that start out very close to each other separate exponentially with time, and in a finite time their separation attains the size of the accessible state space... In the limit of infinite time the Lyapunov exponent is a global measure of the rate at which nearby trajectories diverge, averaged over the strange attractor. (Mathiesen, 2013, p. 114).

During his studies of nonequilibrium motion, Russian physicist Aleksandr Lyapunov found that, out of randomness in very large number sets, periods of order emerged which are now described as the Central Limit Theorem. While conducting this research,

he also developed a formula known as the Lyapunov exponent [$|\delta x(t)| \approx e^{\lambda t} |\delta x_0|$]. In dynamic systems, the Lyapunov exponent suggest a visual representation of iterative chaotic trajectories over time in phase spaces, subscribing to the underlying epistemology of Prigogine's (1996) *arrow of time*. In particular, the Lyapunov exponent helps to explain attractor patterns that are not linear in nature such as those described in detail in "Strange attractors and human interaction" (Gilstrap, 2005). Bertuglia and Vaio (2005) also provide a novel example of Lyapunov instability in relation to Malthus' (1789/1999) famous population studies. In a system where we see large population

growth over time, the incorporation of the Lyapunov exponent into the formula reveals that growth segments in phase spaces are iterative, yet the influence of perturbations makes the spaces between perturbations grow exponentially over time (figure 6). Applied to studies of educational research, this might provide a methodological framework for understanding chaotic episodes in socio-politico-economical phenomenon over time, such as the emergence and growth of charter schools, negative correlations between standardized testing in math and science with reading and spelling, or the like.

When viewing Lyapunov instability, we see evidence of system dynamics responding to initial conditions. For each perturbation, there is an unstable trajectory of exponential divergence measured exponentially against the original system state. As a result, "Lyapunov exponents... provide important quantitative information on the chaotic properties of the evolution of a dynamical system" (Bertuglia & Vaio, 2005, p. 178). Translated into the theoretical constructs of attractor patterns in chaotic systems (Gilstrap, 2005), the following propositions emerge:

1. If all the Lyapunov exponents are negative, then we have a fixed point that behaves like an attractor. Each negative Lyapunov exponent provides a measurement of the speed with which the corresponding state variable directs itself towards the point attractor.
2. If some exponents are nil and others are negative, then we have an ordinary attractor that is less simple than a fixed point, such as for example a limit cycle.
3. If at least one of the Lyapunov exponents... is positive, **then we are in the presence of evolutive dynamics which... diverge exponentially, and give rise to chaos.** Vice versa, we can also state that in a chaotic system at least one Lyapunov exponent is positive. (Bertuglia & Vaio, 2005, pp. 174-175).

However, Bertuglia and Vaio (2005, p. 178) caution that, if we increase the amount of variables we add into the complex model to a large degree, "we do not have the formal methods to solve a system of equations (i.e. to make a model evolve) in exact terms, the formally 'exact' theoretical orbit, basically, is nothing more than a completely abstract concept."

One of the important aspects of Lyapunov exponents in studies in educational settings pertains to degrees of freedom in time series analyses. Davis (2004) has emphasized, "a fractal, for instance, is *scale independent*, meaning that its bumpiness of detail remains constant no matter how much it is magnified or reduced." When we discuss the importance of scale independence in fractal emergence, a similar phenomenon exists in computing Lyapunov exponents. Unlike conservative systems where degrees of freedom are tightly controlled within the quantitative model that is selected for analysis, chaotic systems exhibit different degrees of freedom for each scale of observation and analysis. As Bar-Yam (1997) points out, in dissipative systems information is exchanged between each one of these scales which influences the complexity and subsequent number of Lyapunov exponents of a system. When this happens, "for such systems, we cannot describe the past from present information on a particular length scale (p. 736)," and the complexity of the Lyapunov exponents "can be time dependent, as it is in many irreversible processes (p. 738)."

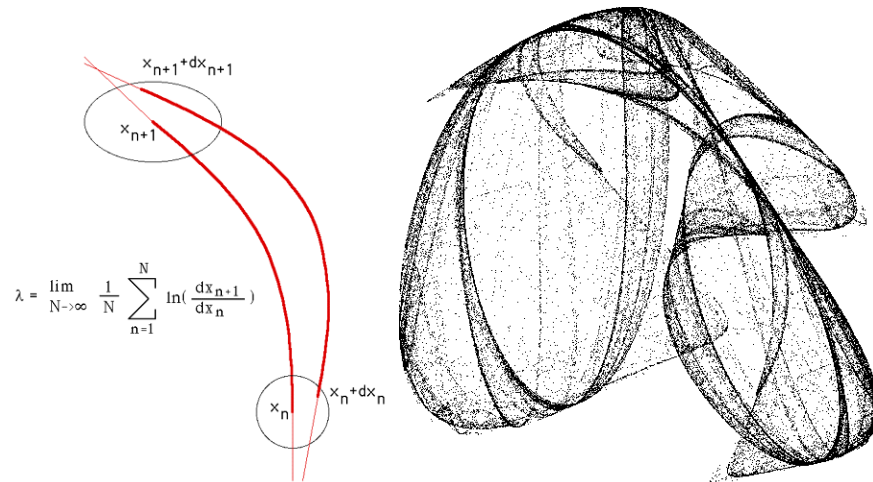


Figure 6: Lyapunov Exponent and Graphical Representation

A *post hoc* postulate of quantitative methods for complexity research: Multicollinearity

In linear regression models, the concept of multicollinearity is one some debate with seemingly no resolution. Multicollinearity can be defined as:

“The situation in which two or more predictors (or subsets of predictors) are strongly (but not perfectly) correlated to one other, making it difficult to interpret the strength of the effect of each predictor (or predictor subset)” (NRG, 2003).

In a nutshell, multicollinearity is determined by computing *eigenvalues* within the variable matrix that holds independent variables in constant with one another to assess the strength of each variable’s ability to predict the dependent variable. An eigenvalue itself is merely the outcome of computing a dimensional matrix of Cartesian coordinates, known as an eigenvector. In the perfect Multiple Linear Regression model, all eigenvalues are calculated at 1. However, if several eigenvalues diverge from 1, multicollinearity is said to exist. In many statistical software packages, this eigenvector is known as the Condition Index.

It is standard practice among many social and behavioral sciences researchers to deal with multicollinearity in a very cutthroat manner: when looking at the matrix of variables, those that show interaction effects (typically ≥ 0.50) are simply discarded. The logic behind this is that the variables have not been properly constructed and defined *a priori*, and that is why the interaction effects exist. As an example, one could argue that in a political research poll where political party affiliation showed interaction effects, the stronger variable chosen for the model should have been whether someone was active in politics in general. This might hold true for a particular study, but when issues such as gender, race, and socio-economic status further promote interaction effects among party affiliation, the logic behind variable removal begins to fall apart. If a decision to remove variables from a model – some of which are critical to understanding – is based on the computation of an eigenvalue, then ultimately the researcher is *letting the software make the decision for the researcher*. It is as if the condition index of the collinearity matrix is being used as an omnibus test for rejection of variables when well known quantitative methods researchers have stated the opposite. As examples, Fox (1991), Cohen and Cohen (1983), Jacard and Turrisi (2003), and Licht (1998) have each turned the researcher toward descriptive analysis in the presence of multicollinearity. And it is this *descriptive* framework that emerges from quantitative models that lends

itself well to the methods of chaos and complexity theories. Unfortunately, the practice of describing variables and the subsequent phenomena that emerge at that point is rarely executed in quantitative studies when multicollinearity occurs, with the practice of variable removal being not only common but almost paramount in positivist research.

So let's walk through exactly what we are discussing. In figure 7, we see evidence of multicollinearity taking place at a few dimensions of a regression model introduced from this author's previous research (Gilstrap & Dupree, 2008a, 2008b). Each of these independent variables [VAR] represents items such as gender, age, year in school, GPA, time period of instruction, semester of instruction, instructor, etc., tested against the dependent variable *scale of reflective thinking*. As referenced previously, eigenvector calculations lead to the presentation of the Condition Index; in this case the value is rather high (CI=30.325, see figure 7). Interaction effects appear to be taking place between [variable] VAR6 and VAR2, VAR7 and VAR1, VAR9 and VAR7, and VAR9 and VAR8 when held constant against the dependent variable. However, correlation analysis showed there were no correlations between these variables without the involvement of the dependent variable, or one of the independent variables is acting as a moderator variable, as discussed previously in this article (Aguinis, 2004). How can this be when some positivist methods tell us the opposite should take place? As stated previously, the condition index is not an omnibus test for rejection, rather here is a time when the researcher should move into descriptive analysis of the phenomenon that are emerging during the model's calculations.

We then performed descriptive statistic analysis for each of the variables that exhibited interaction effects by viewing the mean scores of each variable collectively as a complex system. In effect, this author introduced chaotic and complex systems quantitative analysis as soon as interaction effects were observed. *This created an iterative and recursive method of observation that simultaneously led to emergent post hoc methods of analysis at each stage and level of description.* Rather than choosing the macro level of description as a reductionist and generalizable framework for understanding, this author started with the micro level of description to provide an emergent method of analysis. By understanding that these variables formed a sub-group, which previously was not accounted for, phenomena did indeed emerge that could not be explained by the reductionist model. This process revealed that there are indeed sizable differences in mean scores when scaled at nominal, ordinal, and continuous levels. Furthermore, the more robust level of description of micro analysis for each of these variables and their interactions, the richer the level of emergent description each interacting variable and subgroup of variables added to a more meaningful representation of the macro level of observation and understanding. As an example, time of day was not a statistically significant variable, whereas GPA was significant. However, students who participated in the teaching/learning process during afternoon classes had lower mean scores regardless of GPA where an interaction effect was taking place, such as the interaction between time of day and year in school. Equally, qualitative methods were also conducted to see if observations from quantitative methods could contribute to the qualitative understanding of different emergent phenomena (Gilstrap & Dupree, 2008a, 2008b). By returning to narratives provided by students, therefore, a much richer understanding of these interaction effects began to emerge.

As a result, much of what Bertuglia & Vaio (2005) were arguing theoretically was actually emerging in this applied research. Take, if you will, the number of descriptive and qualitative observations Charles Darwin observed over a period of many years, which led to the formulation of *The Origin of Species*. Had Darwin entered into his research with all of his variables determined and constructed *a priori*, the anomalies he encountered would never have been considered. More importantly, these very anomalies were at the crux of his theory of evolution. Had they been removed from a positivist model of linearity in favor of more robust variables, his findings would have

revealed little more than that the species of the Galapagos Islands did not fit into a generalizable model and were therefore unimportant in understanding life. And as we now know, this is simply untrue.

Collinearity Diagnostics(a)														
Model	Dimension	Eigenvalue	Condition Index	Variance Proportions										
				(Constant)	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10
1	1	8.762	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.950	3.037	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.75	0.00	0.00
	3	0.582	3.879	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.90	0.02	0.00	0.00
	4	0.240	6.047	0.00	0.00	0.01	0.02	0.13	0.21	0.05	0.18	0.00	0.01	0.01
	5	0.115	8.710	0.00	0.23	0.12	0.00	0.08	0.04	0.00	0.01	0.03	0.02	0.19
	6	0.093	9.718	0.00	0.06	0.26	0.02	0.26	0.14	0.00	0.00	0.00	0.19	0.03
	7	0.074	10.867	0.00	0.05	0.26	0.11	0.18	0.12	0.00	0.00	0.03	0.10	0.23
	8	0.070	11.183	0.00	0.22	0.20	0.01	0.00	0.01	0.00	0.01	0.05	0.57	0.08
	9	0.057	12.419	0.00	0.17	0.02	0.17	0.19	0.21	0.00	0.00	0.25	0.03	0.26
	10	0.047	13.605	0.00	0.04	0.00	0.29	0.04	0.19	0.02	0.00	0.00	0.55	0.06
	11	0.010	30.325	0.99	0.21	0.12	0.39	0.11	0.08	0.00	0.02	0.08	0.03	0.02

Figure 7: Dimension matrix of multicollinearity

Conclusion

For the purposes of this article, I have explored the power of quantitative chaos and complexity methods that show potential for further understanding of educational research that involves humans. I then proposed an example of a quantitative method that has *post hoc* implications for chaos and complexity theorists. In a model such as multiple regression, what do we do when the data becomes messy? A reductionist might say: “simply reduce the amount of variables to make the model work cleanly”. A complexivist might contend: “now wait a minute, there is something happening here, let’s describe it”. It is true that interaction effects between variables can be the result of poorly defined variables at the beginning of a research project (e.g., two variables are very similar in nature). However, there is something unique taking place when we see interaction effects between variables that have little similarity. In essence, we have discovered serendipitously a quasi-correlation between two entities that would otherwise never have been noticed; a dynamic has emerged within a system that a reductionist might have written off all together. Moreover, the concept of scale independence equally emerges in this example, since: 1) it is only when these independent variables are held constant against each other and simultaneously with the dependent variable that interactions emerge; 2) understanding these interactions cannot be undertaken through a generalized model, rather each interaction must be explored in a descriptive framework; 3) analysis of means when scaled at a more comprehensive level reveals how and why interactions might be taking place; and 4) understanding how variables influence each other at the micro level leads to an emergent and subsequently richer understanding of the macro level of the research study. In other words, emergence, rather than reduction, becomes paramount in describing the phenomenon revealed *post hoc* of the original model.

Although we are sometime reluctant to suggest quantitative methods with our students when chaotic and/or complex phenomenon emerges, situations such as these might enable us to bridge some of the divides between qualitative and quantitative research. Rather than stopping at the point when the data becomes messy, perhaps we might step back and question whether something larger and more complex is taking place. If we feel it might be, we should explore *post hoc* analyses to determine if chaotic episodes could be emerging. A situation such as in the proposed example might be a

particular case for this type of thinking. In effect, we have taken what appears to be a benign and typical regression approach and have explored it in more depth, revealing the emergence of additional phenomena. Continuing from the micro to the macro levels of this particular system, what opportunities do we then have to explore these phenomena through quantitative methods such as Monte Carlo methods, Markov chains, and Lyapunov exponents? Do anomalies such as this example reinforce the need for a better understanding of the role and purpose of time in our research while equally challenging reversibility as a universal constant? Are methods such as these appropriate, and do they add new levels of analyses, which might actually lead to more robust methods for conducting research on chaos and complexity phenomena in the future? It is my hope that this dialogue will continue among scholars of chaos and complexity theories in educational research.

References:

- Abraham, F. D., & Gilgen, A. R. (Eds.). (1995). *Chaos theory in psychology*. Westport, CT: Greenwood Press.
- Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York: The Guilford Press.
- Bar-Yam, Y. (1997). *Dynamics of complex systems*. Reading, MA: Addison-Wesley.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Bertuglia, C. S., & Vaio, F. (2005). *Nonlinearity, chaos, and complexity: The dynamics of natural and social systems*. Oxford, UK: Oxford UP.
- Boccaro, N. (2004). *Modeling complex systems*. New York: Springer-Verlag.
- Bütz, M. R. (1997). *Chaos and complexity: Implications for theory and practice*. Washington, D. C.: Taylor & Francis.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.)*. London: Lawrence Erlbaum Associates.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper-Collins.
- Davis, B. (2004). *Inventions of teaching: A genealogy*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Fox, J. (1991). *Regression diagnostics*. London: Sage Publications.
- Gilstrap, D. L. (2005). Strange attractors and human interaction: Leading complex organizations through the use of metaphors. *Complicity: An International Journal of Complexity and Education*, 2(1), 73-86.
- Gilstrap, D. L., & Dupree, J. (2008a). A regression model of predictor variables on critical reflection in the classroom: Integration of the Critical Incident Questionnaire and the Framework for Reflective Thinking. *Journal of Academic Librarianship*, 34(6), 469-481.
- Gilstrap, D. L., & Dupree, J. (2008b). Assessing learning, critical reflection, and quality educational outcomes: The Critical Incident Questionnaire. *College & Research Libraries*, 69(5), 407-426.
- Gilstrap, D. L. (2010). Critical reflection as an irreversible process: Epicurus, the Arrow of Time, and an ontology for organizational learning phenomena. *Emergence: Complexity and Organization*, 12(4), 95-117.
- Gleick, J. (1987). *Chaos: Making a new science*. New York: NY: Penguin.
- Grinstead, C. & Snell, J. L. (1997). *Introduction to probability*. American Mathematical Society.
- Green, P. J., & Silverman, B. W. (2000). *Nonparametric regression and generalized linear models*. New York: CRC Press, Taylor & Francis Group.
- Gregson, R. A. M. (2010). Markov chain example: Transitions between two pictorial attractors. In S. J. Guastello & A. M. Gregson (Eds.) *Nonlinear dynamical systems analysis for the behavioral sciences using real data* (pp. 425-444). New York: CRC Press, Taylor & Francis Group.
- Guastello, S. J. (2011). Nonlinear regression and structural equations. In S. J. Guastello & R. A. M. Gregson (Eds.), *Nonlinear dynamical systems analysis for the behavioral sciences using real data* (pp. 281-304). New York: CRC Press, Taylor & Francis Group.
- Guastello, S. J., Johnson, E. A., & Rieke, M. L. (1999). Nonlinear dynamics of motivational flow. *Nonlinear Dynamics, Psychology, and Life Sciences*, 3, 259-274.
- Guegan, D. (2001). Some remarks on the statistical modeling of chaotic systems. In A. I. Mees (Ed.), *Nonlinear dynamics and statistics* (pp. 105-126). Boston: Birkhäuser.
- Hayes, B. (2013). First link in the Markov Chain. *American Scientist*, 111, 92-97.

- Hesse-Biber, S. N., & Leavy, P. (Eds.). (2008). *Handbook of emergent methods*. New York: The Guilford Press.
- Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression (2nd ed.)*. Thousand Oaks, CA: Sage.
- Lemieux, C. (2009). *Monte Carlo and quasi-Monte Carlo sampling*. New York: Springer.
- Liang, F., Liu, C., & Carroll, R. J. (2010). *Advanced Markov chain Monte Carlo methods: Learning from past samples*. West Sussex, UK: John Wiley & Sons.
- Licht, M. H. (1998). Multiple regression and correlation. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 19-64). Washington, DC: American Psychological Association.
- Malthus, R. (1798/1999). *An essay on the principle of population*. Oxford, UK: Oxford University press.
- Mathiesen, J. (2013). Lyapunov exponents. In P. Cvitanović, R. Artuso, P. Dahlqvist, R. Mainieri, G. Tanner, G. Vattay, N. Whelan & A. Wirzba (Eds.), *Chaos: Classical and quantum* (pp. 114-124). Copenhagen, Denmark: Niels Bohr Institute.
- Merrill, S. (2010). Markov chains for identifying nonlinear dynamics. In S. J. Guastello and A. M. Gregson (Eds.), *Nonlinear dynamical systems analysis for the behavioral sciences using real data* (pp. 401-423). New York: CRC Press, Taylor & Francis Group. .
- Metropolis, N. & Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association*, 44(247), pp. 335-341.
- NRG: *Nature Reviews Genetics* (2004). "Markov Chain." 5(4): 1318. Accessed February 1, 2011, at: http://www.nature.com/nrg/journal/v5/n4/glossary/nrg1318_glossary.html.
- NRG: *Nature Reviews Genetics* (2003). "Multicollinearity." 4(2): 996. Accessed February 1, 2011, at: http://www.nature.com/nrg/journal/v4/n2/glossary/nrg996_glossary.html.
- Poincaré, H. (1892/1987). *Les méthodes nouvelle de la mécanique céleste*. Paris: Gauthier-Villars et Fils.
- Prigogine, I. (1996). *The end of certainty: Time, chaos, and the new laws of nature*. New York: The Free Press.
- Prigogine, I., & Stengers, I. (1984). *Order out of chaos: Man's new dialogue with nature*. New York: Bantam Books, Incorporated.
- Stark, J. (2001). Delay reconstruction: Dynamics versus statistics. In A. I. Mees (Ed.) *Nonlinear Dynamics and Statistics*. Boston: Birkhäuser, pp. 81-103.
- Waldrop, M. M. (1992). *Complexity: The emerging science at the edge of order and chaos*. New York: Simon & Schuster.

Figures:

Figure 1: Taken from researcher's SPSS MLR Analysis

Figure 2: <http://www.analyticbridge.com/forum/topics/2004291:Topic:23301>

Figure 3:

http://orion.uwaterloo.ca/~hwoikowi/henry/reports/talks.d/t09talks.d/09waterloomatlab.d/optimTipsWebinar/html/optimTipsTricksWalkthrough_10.png

Figure 4: http://www.clear.rice.edu/comp130/12spring/bayes/monte_carlo_air_pollution_lg.gif

Figure 5: <http://cnx.org/content/m34949/latest/MarkovChainCropped.png>

Figure 6: <http://paulbourke.net/fractals/lyapunov/>

Figure 7: Taken from researcher's SPSS MLR Analysis

About the Author

Donald Gilstrap is Professor and Dean of University Libraries and Professor of Educational Leadership at Wichita State University, Wichita, KS, USA. He has worked with chaos and complexity theories for a number of years, and his research interests include nonlinear dynamics, leadership, organizational change, research methods, and curriculum theory. He may be reached at donald.gilstrap@wichita.edu