Modeling complex systems, from the growth of cancer cells and bacteria to the evolution of a species, is critical to our understanding of the physical world and our place in it. To comprehend our origins, it is essential to accurately model the dynamics driving the creation of life, and to comprehend our future, it is essential to have accurate predictive models. Such modeling is complicated when the dynamics behind these complex systems is random, or stochastic.

Current models describing biological processes are deterministic, meaning they do not accurately reflect the randomness of complex systems over long periods of time. Dr. Motsch, a professor of applied biology and statistics in the School of Mathematical and Statistical Sciences, and I will study the same complex systems with the use of Brownian motion in order to obtain more accurate predictions of the behavior of stochastic complex systems. Brownian motion, originally studied by the botanist Robert Brown in the 19th century and popularized by Einstein in the early 20th century, describes motion that behaves stochastically. For instance, if a particle suspended in water were acted upon randomly by other particles, its behavior would be described as Brownian. My research with Dr. Motsch will seek to demonstrate that Brownian motion will accurately model stochastically complex systems and allow us to invent frameworks that derive meaningful insights into the future.

In order to quantify the accuracy of our approach, we will be using a tool in mathematics called the Wasserstein Distance, which quantifies the distance between functions or data sets, and thereby provides us insight into the improvement that our approach to modeling stochastic systems offers. We aim to collaborate with Dr. Pedro Lowenstein, a professor of Neurosurgery at the University of Michigan to obtain data on cancer growth in a mouse, and with this data, we will use the Wasserstein Distance to compute the distance between the true biological data, the predictions made by the Brownian motion approach, and the predictions made by the deterministic approach. We seek to understand how the number of particles used in the Brownian motion approach affects the accuracy of our approximation of the biological data. More specifically, we will propose an algorithm that will determine the number of particles necessary to get within a desired error using the Brownian motion approach.

During a summer research project, completed under Dr. Anne Gelb in 2014, I studied the use of statistical algorithms to improve Fourier edge detection in the presence of corrupted data. Specifically, Dr. Gelb and I derived an algorithm for reconstructing MRI images that is 166% more accurate than previous methods available to the medical community. By working with Dr. Gelb, I developed an interest in using statistical tools to study relevant problems in biology. During the summer of 2016 I participated in two summer research programs that provided me with the interest and background necessary to successfully complete my Origins Project research. My first project was at the University of California, Santa Barbara where I took a series of courses on differential equations in random media, biological modeling of living systems, and statistical approximations of biological processes. Through conversations with professors and graduate students from Princeton, Stanford, UC Berkeley and others, I learned about cutting-edge research at the intersection of biology and statistics and will apply these techniques in my Origins Project research.

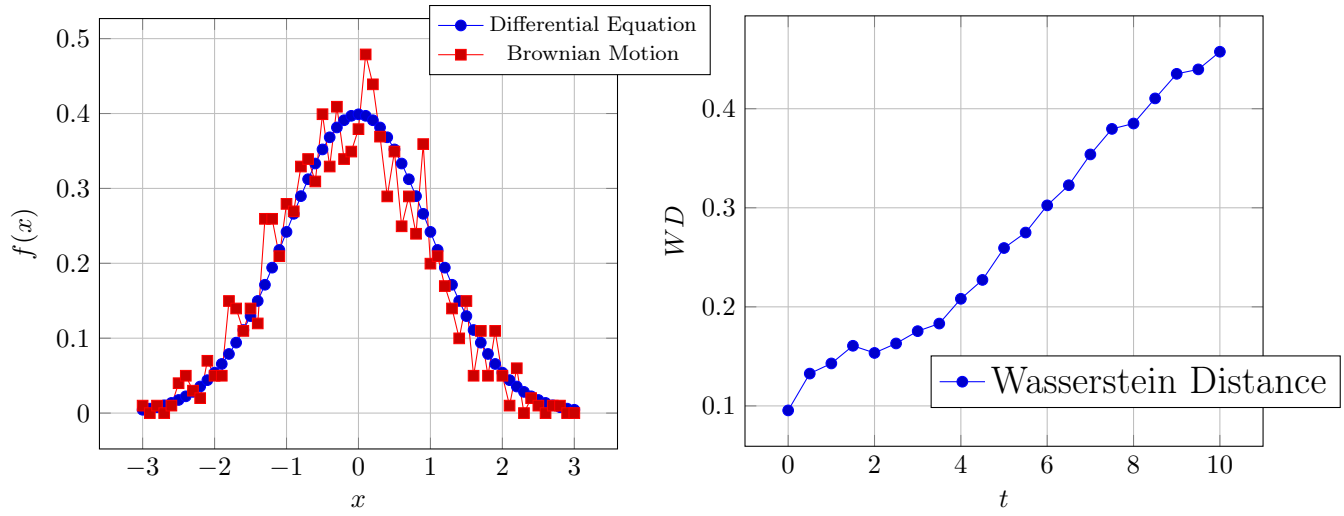
In my second project, I was selected as one of three from over 300 undergraduates to participate in a research experience at San Diego State University. There I studied the convergence behavior of statistical algorithms used in linear regression models. Specifically, my advisor and I proposed the first algorithm for concrete convergence bounds for the error of a Gibbs sampler in a Bayesian linear regression model. These results will be submitted for publication

in October. This research experience taught me the tools to complete statistical convergence analysis, and I will use similar tools to study the convergence behavior of the approximation to complex systems using Brownian motion.

To successfully complete this research, I will need an advanced grasp of real analysis, probability theory, and numerical analysis. . I have taken a course on numerical methods with Dr. Motsch, and I have enrolled in graduate courses in real analysis, distribution theory, and stochastic processes. In these courses I will obtain the tools necessary to analyze papers relating to my work with Dr Motsch, as well as make meaningful contributions to the field. In the Spring I will finish the graduate sequence in real analysis and probability theory, creating a solid theoretical foundation that I will use in my research.

Developing accurate statistical algorithms for understanding stochastic complex systems is one of the largest problems in modern biological sciences, and the opportunity to study this type of problem has been my main focus during my undergraduate career. In addition to contributing to the field of computational biology, I hope to motivate theoretical mathematicians and statisticians to solve problems of interest in applied fields, leading to more interactions and discussions between the previously distant groups of researchers. Indeed, the Wasserstein distance might also be successfully applied to problems in computer science and chemistry. In addition to significantly adding to my knowledge in statistics and computational mathematics, I hope this research project will start a larger discourse on the use of statistical algorithms, specifically Brownian motion, in biology, chemistry, and other applied fields.

The Origins Project represents an avenue towards galvanizing an interest among pure mathematicians and statisticians in using theoretical tools to study problems of interest in biology. With the Origins Project's Undergraduate Research Scholarship, I will have the opportunity to demonstrate the utility of mathematics and statistics in computational biology. By improving the accuracy of approximations in biology, which studies the development of infectious diseases and our place in the universe, my research will shed light on the origin and future of humanity.



(a) Comparing Approaches

(b) Computing the WD

Figure 1

Modeling complex systems, from the growth of cancer cells and bacteria to the movement of stars and planets, is critical to our understanding of the physical world and our place in it. Scientists studying the spread of infectious diseases must understand the virus's behavior to effectively treat patients, and by accurately modeling evolutionary dynamics we are able to grasp our origins, as well as possibily our future. Modeling these processes becomes difficult when the system is stochastic, and new approaches are needed to analyze such models. Brownian motion is an active field of research in statistics, but little research has been done to apply such theory to complex systems. My Origins Project research will derive accurate, computationally efficient algorithms for modeling random complex systems using Brownian motion.

The standard approach to modeling complex systems is to use a differential equation. For instance, a biologist interested in studying cancer growth would try to numerically solve

$$\frac{\partial f}{\partial t} = \left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right) + f(1-f). \tag{1}$$

While this approach is common, it fails to accurately model the stochastic components of the systems. Dr. Motsch and I will examine this same system through the use of Brownian motion. Brownian motion, originally studied by the botanist Robert Brown in the 19th century and popularized by Einstein in the early 20th century, describes motion that behaves in a random way. If an variable's motion is Brownian, its derivative is a normally distributed random variable. By letting X^m denote the particle's location at time m, one can use the following equation to compute its future location

$$X^{m+1} = X^m + C\mathbf{W}, \tag{2}$$

where C is a constant and $\mathbf{W} \sim \mathcal{N}(0,1)$. Substantial research has been done on theoretical Brownian motion, with little progress on applying these results to complex systems.

The goal of my research under Dr. Motsch will be to demonstrate that modeling complex systems with Brownian motion, rather than with a differential equation, leads to a more accurate approximation of the underlying dynamics.

We will study a reaction-diffusion model in one and two dimensions that can be used to study the growth and death of cancer cells in an organism. In this model, cells are created and destroyed randomly, and so we believe modeling this system using a deterministic differential equation leads to incaccurate solutions. This model on its own is used frequently in academia and industry to understand the behavior of cancer and infectious diseases, and thus an improved solution to this important model will have substantial impact on theoretical and computational biology. To show that the Brownian motion approach is more accutarate, we will compare our results to real-world data obtained from an ASU biology lab. In order to make these comparisons, we will be using a tool in pure mathematics called the Wasserstein Distance (WD), which measures the difference between two data sets. The WD will enable us to quantify the improvement of the Brownian motion approach.

Dr. Motsch and I are well positioned to complete this project. I am an undergraduate student working in applied mathematics who intends to pursue a PhD in statistics. In particular I am interested in applying statistical techniques to solving real-world problems. This past summer I was selected along with three other undergraduates out of over 300 to participate in a research program at San Diego State University. There my advisor and I derived the first method of obtaining convergence behavior for the error in algorithms that one uses when estimating coefficients in the linear regression equation. This project resulted in a publication, and I developed an interest in applying statistical tools to study real-world problems. My role in the project will be to appply statistical techniques to modeling Brownian motion in one and two dimensions and applying numerical techniques to solving the differential equations we'll be using. Dr. Motsch received his PhD in applied mathematics with an emphasis in animal cognition and mathematical biology, and works at the intersection of computational math, biology, and statistics. He has overseen undergraduate and graduate theses in flocking behavior, applied statistics, and offers the biological perspective necessary for this project. He has worked previously in the Wasserstein Distance and will provide the background in the WD to apply it to our research.

This project requires an advanced grasp on real analysis, probability theory, and numerical analysis. In order to complete this project successfully, I have enrolled in graduate courses in real analysis, distribution theory, and stochastic processes. In these courses I will obtain the tools necessary to read papers relating to my project as well as hold informed discussions with Dr. Motsch on our research. I have taken a course on numerical methods with Dr. Motsch, and I will use my textbook as a reference during my research. In the Spring I will finish the graduate sequence in real analysis and probability theory.

The Origins Project Scholarship will allow me to focus solely on my research with Dr. Motsch, instead of maintaining an on-campus job to pay for college. This will allow me to make faster progress on my research. Having this research experience will also give me the experience necessary to perform well as a graduate student. We believe that the underlying dynamics of stochastic complex systems are different in nature from deterministic models and are therefore not accurately modeled by differential equations. Through this project I seek to better understand these dynamics through statistical modeling. My research with Dr. Motsch will demonstrate the possibility of using Brownian motion to model stochastic complex systems, shedding light on the random forces that determine the physical world and on our role as a species in it.

Through this research I will gain the necessary research experience and technical background to succesfully enter into a PhD program in statistics. By publishing our results, I will have an advantage in the application process for graduate school. By conducting this research under Dr. Motsch, I'll gain

This project allows me to apply statistical and numerical techniques to solve tangible problems and provides me with an ideal transition to graduate school. We will be using techniques such as Brownian motion, approximations of intractable distributions, finite-difference techniques to study differential equations. lysis and graduate distribution theory in the fall to build upon my theoretical foundation.

Dr. Motsch received his PhD in applied mathematics with an emphasis in animal cognition and mathematical biology, and works at the intersection of computational math and biology. He has mentored several undergraduates working in computational biology, flocking patterns, and cell behavior.

The outline of the project is as follows: In September and October, Dr. Motsch and I will be submitting a paper to journal that demonstrates that agent-based modeling of complex systems accurately represents the behavior of the complex system using data from **Motsch, who is this data from?**. In November and December, we will study and finally in the Spring we will examine.

As a recipient of the Origins Project Scholarship, I will have the chance to devote myself to my research project, rather than working to pay for college, allowing me and Dr. Motsch to make quick progress in our research.

The Origins Project Scholarship will allow me to focus solely on my research with Dr. Motsch, instead of working on campus, allowing for more substantial research progress during the academic year. This will in turn allow me to prepare a stronger graduate school application, The Origins network will provide me and Dr. Motsch with an invaluable network of researchers in biology, chemistry and other fields we can speak with, collaborate with, and learn from. This will in turn lead to research that understands the needs of these fields, leading to more helpful and effective results.

Notes

In Figure 1, the solutions from (1) and (2) are plotted.

Agent is small, the macro might not be realistic, might not describe. send code and kernel, image I have, and send draft.

This approach is computationally faster and places no restriction on the physical and temportal discretization, yet there is a problem. There has been little research done that computes the accuracy of this second scheme in relation to the first. Can one trust the answers given by the second approach?

The goal of my and Dr. Motsch's research is to prove that the second approach provides an accurate approximation to the system one is considering. We do this using both numerical and pure mathematics. What we show is that researchers in fields as diverse as biology, chemistry, medecine, can faithfully rely on the approximations coming from the second approach. The end result is that researchers in these fields can now solve probems that once seemed inaccesible a few years ago because essentially their computing power has been increased significnatly.

The answer is that before now, there was very little way to compare the accuracy of the second approach to the first approach. Some ways existed, but they were complicated and required advanced mathematics to fully understand. What Dr. Motsch and I are doing is simplifying the mathematics and deriving simple algorithms that researchers in fields as diverse as biology, chemistry, medecine, etc. can use in their research. What we are doing is allowing these researchers to avoid the complicated mathematics and use the second approach, knowing that this approach is "close enough" to the first approach to not be sacrificing accuracy.

Our first paper is set to be submitted in October. This paper demonstrates that it is possible to use this second approach in approximating biological processes that can be approximated through differential equations. This means that most problems in biology can now be solved through the second approach. Our research indicates that the second approach is 4 times faster than the first approach.

This project lies at the intersection of evolutionary and computational biology, pure mathematics, applied mathematics, and statistics. Dr. Motsch and I are qualified to carry out this project for several reasons. Dr. Motsch received his PhD in applied mathematics with an emphasis in animal cognition and mathematical biology, while I am undergraduate student in pure mathematics intending to obtain a PhD in statistics. Having the opportunity to conduct research as a Origins Scholarship Recipient will allow me to focus solely on my research and finish a second paper with Dr. Motsch. Having three papers on my CV when I apply to PhD programs will be incredibly beneficial.

The timeline for our project is as follows. In August and September, Dr. Motsch and I are completing work on our first paper, which examines some appliations of the Wasserstein distance to studying the growth of cancer cells. This paper will be submitted at the beginning of October. We will then begin working on a second paper,

there is a miscmatch, we sometimess need to use micro because it more accurately reflects the underlying truth. agent-based models, complex systems.

Novel Techniques to Modeling Complex Systems

Comparison of Agent-based and agent-based models for complex systems:

Modeling Complex Systems: Agent-based Models vs. Macro Models. Contrasting

you need large particle for consistency, we need to shift that the macro works if number is large enough, but sometimes the system is too small so that the validity of the macro is questionable.

it simplifies the computation and allows for more diverse models to be considered.

two figures: one where it works, one where it doesnt work., the one where that it doesnt work, in some sort of experimental data, we'll hold off for now. we want to use opacity, level curves, difference both macro and micro.

open question: compare micro vs. macro, sometimes one doesnt work when n isnt large enough. We need statistics to study the behavior of agent-based when the macro is not enough.

To solve such an equation, one discretizes the physical space (corresponding to the x and y variables) and the temporal space (corresponding to the t variable). This approach requires the ratio $\frac{\Delta t}{\Delta x^2}$ to be very small in order for the solution to be accurate. This results in a computationally expensive problem. Furthermore, as the physical space gets larger, the computational time required increases exponentially. For problems of even moderate size, this approach is therefore impracticle.

Many biological processes can be

As our computing power inceases, questions that once seemed impossible to analyze become accessible. Faster computation has faciliatated new discoveries in computer science, physics, and chemistry, but nowhere are these discoveries more important to our understanding of the physical world than in biology. Biological processes, such as the growth of cancer, the spread of disease, or the evolution of a species, are at the forefront of modern science, and understanding these processes is critical to understanding our place in the universe. Working under Dr. Sébastien Motsch, my Origins Project work will focus on deriving faster, more accurate approaches to analyzing complex dynamical systems.