



REPUBLIC OF THE PHILIPPINES
SORSGON STATE UNIVERSITY
BULAN CAMPUS

OVERVIEW OF COMPUTATIONAL SCIENCE AND THE MODELING PROCESS

CS 313
INTRODUCTION TO
COMPUTATIONAL SCIENCE
1st Semester, S.Y.2021 - 2022
BSCS III

Engr. Rey C. Rodriguez, PhD(Cand.)
Associate Professor II

1 Introduction

Computational science, the fast-growing interdisciplinary field that is at the intersection of the sciences, computer science, and mathematics, will require scientists who are appropriately trained. Experts predicted that future scientists who are not computationally and mathematically literate will be unable to do science [1, 2]. Industry and many other aspects of the commercial world use simulation and computation in diverse ways.

This module gives an overview about computational science and the modeling process. Several applications of computational science are presented, and the modeling process is explained in detail. Since this area is emerging, to preserve accuracy of information, most of the texts are taken verbatim from [1], our reference book.

2 Objectives

After completing this module, you will be able to:

- Investigate applications of computational science involving different scientific areas,
- Investigate an application of computational science, and
- Compare and contrast the computational science modeling process with the scientific method and software life cycle.

3 Discussion of Content

3.1 Overview of Computational Science

Computational science, which combines computer simulation, scientific visualization, mathematical modeling, computer programming, data structures, networking, database design, symbolic computation, and high-performance computing, can transform practices in a diverse range of disciplines. Its computer models and simulations offer valuable approaches to problems in many areas, as the following examples indicate [1].

1. Scientists at Los Alamos National Laboratory and the University of Minnesota wrote, “Mathematical modeling has impacted our understanding of HIV pathogenesis. Before modeling was brought to bear in a serious manner, AIDS was thought to be a slow disease in which treatment could be delayed until symptoms appeared, and patients were not monitored very aggressively. In the large, multicenter AIDS cohort studies aimed at monitoring the natural history of the disease, blood typically was drawn every six months. There was a poor understanding of the biological processes that were responsible for the observed levels of virus in the blood and the rapidity at which the virus became drug resistant. Modeling, coupled with advances in technology, has changed all of this.” Dynamic modeling has not only revealed important features of HIV pathogenesis but has advanced the drug treatment regime for AIDS patients [3, 1]. Since then, [3] and other researchers have applied modeling to enhance our understanding of the hepatitis C virus, which causes widespread infections and is the primary cause of liver cancer in the United States. Such models have already revealed much about the pathogenesis of the virus, the effectiveness of treatments (interferon/ ribavirin and direct antiviral agents), and the influence of genetic variants in the kinetics of the virus [4].
2. From the 1960s, numerical weather prediction has revolutionized forecasting. “Since then, forecasting has improved side by side with the evolution of computing technology, and advances in computing continue to drive better forecasting as weather researchers develop improved numerical models” (Pittsburgh Supercomputing Center 2001). A Weather Research and Forecasting (WRF) Model was released in 2000. The latest version of this model utilizes “multiple dynamical cores, a 3- dimensional variational data assimilation system, and a software architecture allowing for computational parallelism

and system extensibility.” This sophisticated, mesoscale [horizontal scale of 2 to 2000 kilometers (km)] numerical weather-prediction system is useful for forecasting and research. An array of partners, including the National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (NOAA), other governmental and military organizations, universities, plus some international groups, continuously revise the WRF model. With this efficient, adaptable model for forecasting, researchers can conduct simulations using real data or idealized designs. In 2007, a specialized version of WRF was initiated for forecasting and research on hurricanes (WRF)[5, 1].

3. A multidisciplinary team at the University of Tennessee’s Institute for Environmental Modeling is using computational ecology to study complex options for ecological management of the Everglades. Louis Gross, Director of the Institute, says that “computational technology, coupled with mathematics and ecology, will play an ever-increasing role in generating vital information society needs to make tough decisions about its surroundings” (Lynn 2003). South Florida has a wellknown history of disruptions to normal water flow. The UT group has developed a parallelized landscape population model (ALFISH) to integrate with other models in a multiscale ecological multimodel (Across-Trophic Level System Simulation, ATLSS). ALFISH is used to model the effects on freshwater fish (planktivorous and piscivorous) of different water-management plans. These fish populations represent food resources for wading birds, and researchers can link the fish model with wading bird models to help sustain the higher-level multimodel [6].
4. Application of computer modeling has fueled the debate in another, rather unexpected area—linguistics. The origin of the Indo-European family of languages is rather hotly debated between proponents of two hypotheses—Eurasian steppes, 6000 years (yr) ago versus Anatolia (mostly in present-day Turkey), 8000 to 9500 years ago. This family of languages has given rise to more than 400 modern languages, spoken by about three billion people. Recently, Quentin Atkinson and colleagues utilized evolutionary models, often employed to ascertain the origin of viruses that lead to epidemics, to analyze this problem. Based on common vocabulary words from various languages in the family, the model supports Anatolian origin, as agricultural techniques were broadcast. Although certainly not resolving the argument, the results have given experts in this

field something to consider carefully [7].

3.2 The Modeling Process

The process of making and testing hypotheses about models and then revising designs or theories has its foundation in the experimental sciences. Similarly, computational scientists use **modeling** to analyze complex, real-world problems in order to predict what might happen with some course of action. Some examples are [1]:

- Professor Muneo Hori and colleagues from the Earthquake Research Institute, the University of Tokyo, Japan, use high-performance computation with sophisticated models to simulate earthquakes, making quantitative predictions of infrastructural damages, response, and recovery to help minimize damage, death, and injury [8].
- Professor Liming Liang, statistical geneticist at Harvard School of Public Health, uses computational and statistical tools to better understand the genetic variation in complex human diseases, such as dyslipidemia, cancer, and type 2 diabetes [9].
- Cognitive scientists, such as Professor Ken Koedinger at the Human-Computer Interaction Institute, Carnegie-Mellon University, develop computer models of student reasoning and learning to aid in the design educational software and to guide teaching practices [10].
- Both civilian and military organizations commonly employ drones (unmanned aerial vehicles, UAVs). Whether to monitor air quality or supervise combat forces, this technology is becoming more and more important, but the operation of the drones is quite complex.
- While a postdoctoral fellow at MIT, Dr. Luca Bertuccelli worked with a team using models to develop new decision support systems, enabling the operators of these craft to make better decisions [11].
- A mathematical modeling team at the Universidad Nacional Autónoma de México has been modeling the dynamics of infections such as West Nile encephalitis, dengue fever, and yellow fever. A better understanding of these viruses will improve outbreak predictions, interventions, and responses [12].

Several classification categories for models exist. A system we are modeling exhibits **probabilistic**, or **stochastic, behavior** if it appears that an element of chance exists. For example, the path of a hurricane

Definition **Modeling** is the application of methods to analyze complex, real-world problems in order to make predictions about what might happen with various actions.

is probabilistic. In contrast, a behavior can be **deterministic**, such as the position of a falling object in a vacuum. Similarly, models can be deterministic or probabilistic. A **probabilistic**, or **stochastic**, **model** exhibits random effects, while a **deterministic model** does not. The results of a deterministic model depend on the initial conditions; and in the case of computer implementation with particular input, the output is the same for each program execution. We can have a probabilistic model for a deterministic situation, such as a model that uses random numbers to estimate the area under a curve.

Definitions A system exhibits **probabilistic**, or **stochastic**, **behavior** if an element of chance exists. Otherwise, the system exhibits **deterministic behavior**. A **probabilistic**, or **stochastic**, **model** exhibits its random effects, while a **deterministic model** does not.

We can also classify models as static or dynamic. In a **static model**, we do not consider time, so that the model is comparable to a snapshot or a map. For example, a model of the weight of a salamander as being proportional to the cube of its length has variables for weight and length but not for time. By contrast, in a **dynamic model**, time changes, so that such a model is comparable to an animated cartoon or a movie. For example, the number of salamanders in an area undergoing development changes with time; hence, a model of such a population is dynamic. Many of the models we consider in this text are dynamic and employ a static component as part of the dynamic model.

Definitions A **static model** does not consider time, while a **dynamic model** changes with time.

When time changes continuously and smoothly, the model is **contin-**

uous. If time changes in incremental steps, the model is **discrete**. A discrete model is analogous to a movie. A sequence of frames moves so quickly that the viewer perceives motion. However, in a live play, the action is continuous. Just as a discrete sequence of movie frames represents the continuous motion of actors, we often develop discrete computer models of continuous situations.

Definitions In a **continuous model**, time changes continuously, while in a **discrete model**, time changes in incremental steps.

STEPS OF THE MODELING PROCESS

The modeling process is cyclic and closely parallels the scientific method and the software life cycle for the development of a major software project. The process is cyclic because at any step we might return to an earlier stage to make revisions and continue the process from that point.

The steps of the modeling process are as follows [1]:

(a) **Analyze the problem.**

We must first study the situation sufficiently to identify the problem precisely and understand its fundamental questions clearly. At this stage, we determine the problem's objective and decide on the problem's classification, such as deterministic or stochastic. Only with a clear, precise problem identification can we translate the problem into mathematical symbols and develop and solve the model.

(b) **Formulate a model.**

In this stage, we design the model, forming an abstraction of the system we are modeling. Some of the tasks of this step are as follows:

(i) **Gather data.**

We collect relevant data to gain information about the system's behavior.

(ii) **Make simplifying assumptions and document them.**

In formulating a model, we should attempt to be as simple as reasonably possible. Thus, we frequently decide to simplify some of the factors and to ignore other factors that do not seem as important. Most problems are entirely too complex to consider every detail, and doing so would only make the model impossible to solve or to run in a reasonable amount of

time on a computer. Moreover, factors often exist that do not appreciably affect outcomes. Besides simplifying factors, we may decide to return to Step 1 to restrict further the problem under investigation.

(iii) **Determine variables and units.**

We must determine and name the variables. An independent variable is the variable on which others depend. In many applications, time is an independent variable. The model will try to explain the dependent variables. For example, in simulating the trajectory of a ball, time is an independent variable; and the height and the horizontal distance from the initial position are dependent variables whose values depend on the time. To simplify the model, we may decide to neglect some variables (such as air resistance), treat certain variables as constants, or aggregate several variables into one. While deciding on the variables, we must also establish their units, such as days as the unit for time.

(iv) **Establish relationships among variables and submodels.**

If possible, we should draw a diagram of the model, breaking it into submodels and indicating relationships among variables. To simplify the model, we may assume that some of the relationships are simpler than they really are. For example, we might assume that two variables are related in a linear manner instead of in a more complex way.

(v) **Determine equations and functions.**

While establishing relationships between variables, we determine equations and functions for these variables. For example, we might decide that two variables are proportional to each other, or we might establish that a known scientific formula or equation applies to the model. Many computational science models involve differential equations, or equations involving a derivative.

(c) **Solve the model.**

This stage implements the model. It is important not to jump to this step before thoroughly understanding the problem and designing the model. Otherwise, we might waste much time, which can be most frustrating. Some of the techniques and tools that the solution might employ are algebra, calculus, graphs, computer programs, and computer packages. Our solution might produce an exact answer or

might simulate the situation. If the model is too complex to solve, we must return to Step 2 to make additional simplifying assumptions or to Step 1 to reformulate the problem.

(d) **Verify and interpret the model's solution.**

Once we have a solution, we should carefully examine the results to make sure that they make sense (verification) and that the solution solves the original problem (validation) and is usable. The process of verification determines if the solution works correctly, while the process of validation establishes whether the system satisfies the problem's requirements. Thus, verification concerns "solving the problem right," and validation concerns "solving the right problem." Testing the solution to see if predictions agree with real data is important for verification. We must be careful to apply our model only in the appropriate ranges for the independent data. For example, our model might be accurate for time periods of a few days but grossly inaccurate when applied to time periods of several years. We should analyze the model's solution to determine its implications. If the model solution shows weaknesses, we should return to Step 1 or 2 to determine if it is feasible to refine the model. If so, we cycle back through the process. Hence, the cyclic modeling process is a trade-off between simplification and refinement. For refinement, we may need to extend the scope of the problem in Step 1. In Step 2, while refining, we often need to reconsider our simplifying assumptions, include more variables, assume more complex relationships among the variables and submodels, and use more sophisticated techniques.

(e) **Report on the model.**

Reporting on a model is important for its utility. Perhaps the scientific report will be written for colleagues at a laboratory or will be presented at a scientific conference. A report contains the following components, which parallel the steps of the modeling process:

(i) **Analysis of the problem**

Usually, assuming that the audience is intelligent but not aware of the situation, we need to describe the circumstances in which the problem arises. Then, we must clearly explain the problem and the objectives of the study.

(ii) **Model design**

The amount of detail with which we explain the model depends on the situation. In a comprehensive technical report, we can incorporate much more detail than in a conference talk. In

either case, we should state the simplifying assumptions and the rationale for employing them. Clearly labeled diagrams of the relationships among variables and submodels are usually very helpful in understanding the model.

(iii) **Model solution**

In this section, we describe the techniques for solving the problem and the solution. We should give as much detail as necessary for the audience to understand the material without becoming mired in technical minutia. For a written report, appendices may contain more detail, such as source code of programs and additional information about the solutions of equations.

(iv) **Results and conclusions**

Our report should include results, interpretations, implications, recommendations, and conclusions of the model's solution. Usually, we present some of the data and results in tables or graphs. Such figures should contain titles, sources, and labels for columns and axes. We may also include suggestions for future work.

(f) **Maintain the model.**

As the model's solution is used, it may be necessary or desirable to make corrections, improvements, or enhancements. In this case, the modeler again cycles through the modeling process to develop a revised solution

Definitions The process of **verification** determines if the solution works correctly, while the process of **validation** establishes if the system satisfies the problem's requirements.

Although we described the modeling process as a sequence or series of steps, we may be developing two or more steps simultaneously. For example, it is advisable to be compiling the report from the beginning. Otherwise, we can forget to mention significant points, such as reasons for making certain simplifying assumptions or for needing particular refinements. Moreover, within modeling teams, individuals or groups frequently work on different submodels simultaneously. Having completed a submodule, a team member might be verifying the submodule while others are still working on solving theirs.

The modeling process is a creative, scientific endeavor. As such, a problem we are modeling usually does not have one correct answer. The problems are complex, and many models provide good, although different, solutions. Thus, modeling is a challenging, open-ended, and exciting venture.

NOTE: Read also the article[13] about Modeling and Simulation here:
https://repository.lib.ncsu.edu/bitstream/handle/1840.4/7445/1997_0002.pdf?sequence=1&isAllowed=y

4 Asynchronous Activities

WEEK 1

1. Investigate three applications of computational science involving different scientific areas and write at least a paragraph on each. List references.
2. Investigate an application of computational science and write a three-page, typed, double-spaced paper on the topic. List references.

WEEK 2

1. Compare and contrast the modeling process with the scientific method: Make observations; formulate a hypothesis; develop a testing method for the hypothesis; collect data for the test; using the data, test the hypothesis; accept or reject the hypothesis.
2. Compare and contrast the modeling process with the software life cycle: analysis, design, implementation, testing, documentation, maintenance.

References

- [1] A. B. Shiflet and G. W. Shiflet, *Introduction to computational science: modeling and simulation for the sciences*. Princeton University Press, 2014.
- [2] S. Emmott and S. Rison, “Towards 2020 science,” *Science in Parliament*, vol. 65, no. 4, pp. 31–33, 2008.
- [3] A. S. Perelson and P. W. Nelson, “Mathematical analysis of hiv-1 dynamics in vivo,” *SIAM review*, vol. 41, no. 1, pp. 3–44, 1999.
- [4] H. Dahari, J. Guedj, A. S. Perelson, and T. J. Layden, “Hepatitis c viral kinetics in the era of direct acting antiviral agents and interleukin-28b,” *Current hepatitis reports*, vol. 10, no. 3, pp. 214–227, 2011.
- [5] The weather research and forecasting model. [Online]. Available: <https://www.mmm.ucar.edu/weather-research-and-forecasting-model>
- [6] D. Wang, M. W. Berry, E. A. Carr, and L. J. Gross, “A parallel fish landscape model for ecosystem modeling,” *Simulation*, vol. 82, no. 7, pp. 451–465, 2006.
- [7] R. Bouckaert, P. Lemey, M. Dunn, S. J. Greenhill, A. V. Alekseyenko, A. J. Drummond, R. D. Gray, M. A. Suchard, and Q. D. Atkinson, “Mapping the origins and expansion of the indo-european language family,” *Science*, vol. 337, no. 6097, pp. 957–960, 2012.
- [8] M. Lalith and M. Hori, “Application of high performance computation for the prediction of urban area earthquake disaster,” 2011.
- [9] Liang, liming. my research, personal webpage. [Online]. Available: <http://www.hsph.harvard.edu/faculty/liming-liang/index.html>
- [10] Koedinger. k. personal webpage. [Online]. Available: <http://www.hcii.cmu.edu/people/faculty/ken-koedinger>
- [11] Bertuccelli, luca f. personal webpage. [Online]. Available: <http://web.mit.edu/lucab/www/publications.html>
- [12] C. Vargas, L. Esteva, and G. Cruz-Pacheco, “Mathematical modelling of arbovirus diseases,” in *2010 7th International Conference on Electrical Engineering Computing Science and Automatic Control*. IEEE, 2010, pp. 205–209.

- [13] A. Maria, “Introduction to modeling and simulation,” in *Proceedings of the 29th conference on Winter simulation*, 1997, pp. 7–13.