

MATH

MODELING

**GETTING
STARTED &
GETTING
SOLUTIONS**



K. M. BLISS
K. R. FOWLER
B. J. GALLUZZO

PUBLISHER

Society for Industrial and Applied Mathematics (SIAM)
3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA
www.siam.org

FUNDING PROVIDED BY

The Moody's Foundation in association with the Moody's
Mega Math Challenge, the National Science Foundation
(NSF), and the Society for Industrial and Applied
Mathematics (SIAM).

AUTHORS

Karen M. Bliss
*Department of Math & Computer Science,
Quinnipiac University, Hamden, CT*

Kathleen R. Fowler
*Department of Math & Computer Science,
Clarkson University, Potsdam, NY*

Benjamin J. Galluzzo
*Department of Mathematics,
Shippensburg University, Shippensburg, PA*

**DESIGN & CONNECTIONS TO
COMMON CORE STATE STANDARDS**

PlusUs
www.plusus.org

PRODUCTION

First Edition 2014
Printed and bound in the United States of America

No part of this guidebook may be reproduced or
stored in an online retrieval system or transmitted in
any form or by any means without the prior written
permission of the publisher. All rights reserved.

CONTENTS

1. INTRODUCTION

2

2. DEFINING THE PROBLEM STATEMENT

10

3. MAKING ASSUMPTIONS

15

4. DEFINING VARIABLES

20

5. BUILDING SOLUTIONS

25

6. ANALYSIS AND MODEL ASSESSMENT

32

7. PUTTING IT ALL TOGETHER

40

APPENDICES & REFERENCE

45

**THE WORLD
AROUND
US IS FILLED
WITH
IMPORTANT,
UNANSWERED
QUESTIONS.**

1. INTRODUCTION

The world around us is filled with important, unanswered questions. What effect will rising sea levels have on the coastal regions of the United States? When will the world's human population surpass 10 billion? How much will it cost to go to college in 10 years? Who will win the next U.S. Presidential election? There are also other phenomena we wish to understand better. Is it possible to study crimes and identify a burglary pattern [1, 10]? What is the best way to move through the rain and not get soaked [7]? How feasible is invisibility cloaking technology [6]? Can we design a brownie pan so the edges do not burn but the center is cooked [2]? Possible answers to these questions are being sought by researchers and students alike. Will they be able to find the answers? Maybe. The only thing one can say with certainty is that any attempt to find a solution requires the use of mathematics, most likely through the creation, application, and refinement of mathematical models.

A mathematical model is a representation of a system or scenario that is used to gain qualitative and/or quantitative understanding of some real-world problems and to predict future behavior. Models are used in a variety of disciplines, such as biology, engineering, computer science, psychology, sociology, and marketing. Because models are abstractions of reality, they can lead to scientific advances, provide the foundation for new discoveries, and help leaders make informed decisions.

This guide is intended for students, teachers, and anyone who wants to learn how to model. The aim of this guide is to demystify the process of how a mathematical model can be built. Building a useful math model does not necessarily require advanced mathematics or significant expertise in any of the fields listed above. It does require a willingness to do some research, brainstorm, and jump right in and try something that may be out of your comfort zone.

1: INTRODUCTION

MATH MODELING VS. WORD PROBLEMS

Modeling problems are entirely different than the types of word problems most of us encountered in math classes. In order to understand the difference between math modeling and word problems, consider the following questions about recycling.

1. The population of Yourtown is 20,000, and 35% of its citizens recycle their plastic water bottles. If each person uses 9 water bottles per week, how many bottles are recycled each week in Yourtown?

2. How much plastic is recycled in Yourtown?

The solution to the first question is straightforward:

$$0.35 \times 20,000 \text{ PEOPLE} \times 9 \frac{\text{BOTTLES}}{\text{PERSON} \times \text{WEEK}} = 63,000 \frac{\text{BOTTLES}}{\text{WEEK}}$$

This type of question might appear in a math textbook to reinforce the concept that we translate the phrase “35% of” to the mathematical computation “0.35 times.” It is an example of what we would call a word problem: the problem explicitly gives you all the information you need. You need only determine the appropriate math computation(s) in order to arrive at the one correct answer. Word problems can be used to help students understand why we might want to study a particular mathematical concept and reinforce important mathematical skills.

The second question is quite different. When you read a question like this, you might be thinking, “I don’t have enough information to answer this question,” and you’re right! This is exactly the key point: we usually don’t have complete information when trying to solve real-world problems. Indeed, such situations demand that we use both mathematics and creativity. When we encounter such situations where we have

incomplete information, we refer to the problem as open-ended. It turns out that mathematical modeling is perfect for open-ended problems. This question, for example, might have been conceived because we saw garbage cans overflowing with water and soda bottles and then wondered how many bottles were actually being thrown out and why they were not being recycled. Modeling allows us to use mathematics to analyze the situation and propose a solution to promote recycling.

In the word problem example above, it is assumed that each person in town uses 9 plastic water bottles per week and that 35% of the 20,000 people recycle their water bottles every time they use one. Are these reasonable assumptions? The number 20,000 is probably an estimate of Yourtown’s population, but where is the other information coming from? Is it likely that every person in the town uses exactly 9 water bottles every week? Is it likely that 35% of people recycle every water

bottle they use while 65% of people never recycle any of their water bottles? Probably not, but maybe this is an average value, based on other data. The first problem doesn't invite us to determine whether the scenario is realistic; it is assumed that we will accept the given information as true and make the appropriate computations.

In order to answer the second (modeling) problem above, you would need to research the situation for yourself, making your own (reasonable) assumptions and strategies for answering the question. The question statement doesn't provide specific details about Yourtown.

You will have to determine what factors about Yourtown contribute to the amount of plastic that gets recycled. It seems reasonable to believe that the population of Yourtown is an important factor, but what else about the city affects the recycling rate? The question statement failed to mention what types of plastic you should be taking into account. It would be hard to quantify all plastic thrown away. Is it a reasonable assumption to consider only the plastics from food and beverage containers if you believe those are the primary plastic waste sources? You would have to do some research and make some assumptions in order to make any progress on this problem.

If, after your research, you distill the original problem into something very specific, such as "Determine

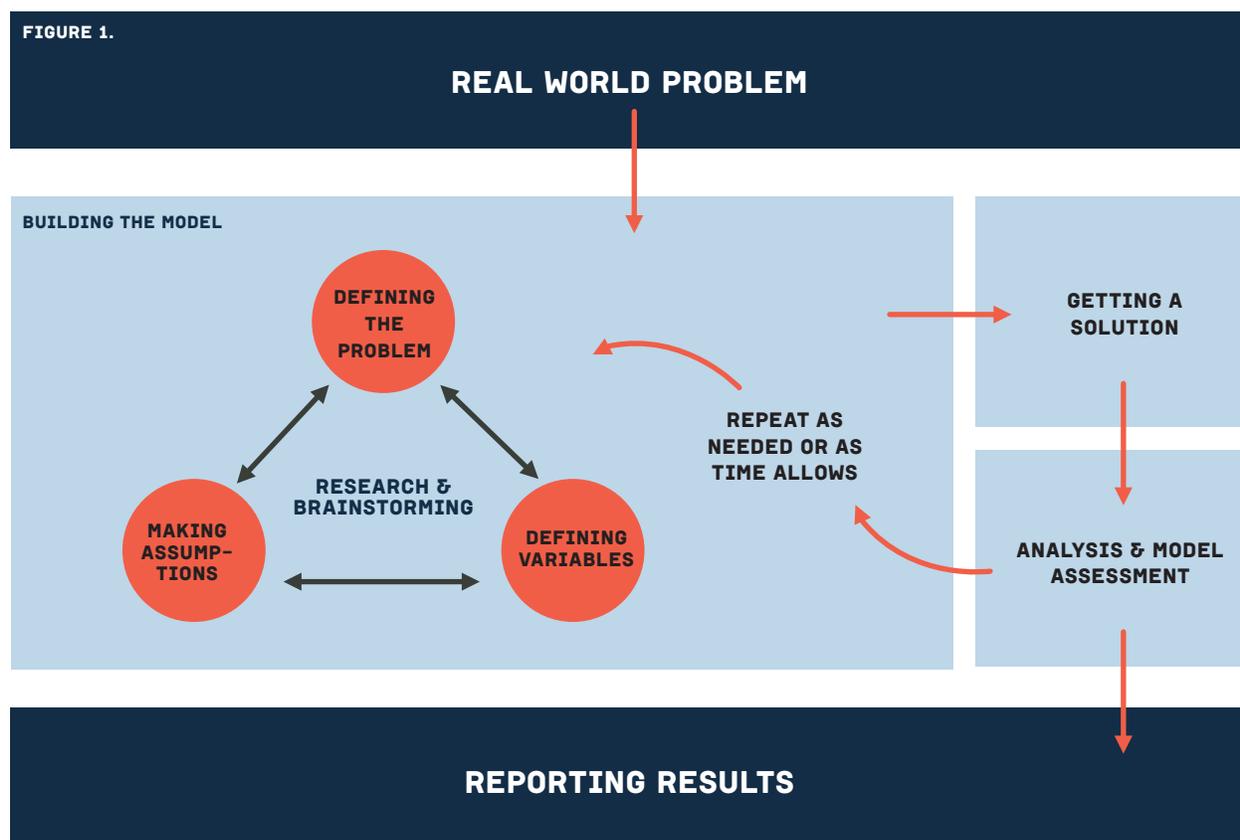
the volume of plastic waste Yourtown sent to landfills last year," then there is exactly one correct answer. However, it's unlikely that you will ever have sufficient information to find that answer. In light of this, you will develop a model that best estimates the answer given the available information. Since no one knows the true answer to the question, your model is at least as important as the answer itself, as is your ability to explain your model.

In contrast to word problems, we often use the phrase "a solution" (as opposed to "the solution") when we talk about modeling problems. This is because people who look at the same modeling problem may have different perspectives into its resolution and can certainly come up with different, valid alternative solutions. It is worth noting that word problems can actually be thought of as former modeling problems. That is to say, someone has already determined a simple model and provided you with all the relevant pieces of information. This is very different from a modeling problem, in which you must decide what's important and how to piece it all together.

Mathematical modeling questions allow you to research real-world problems, using your discoveries to create new knowledge. Your creativity and how you think about this problem are both highly valuable in finding a solution to a modeling question. This is part of what makes modeling so interesting and fun!

PEOPLE WHO LOOK AT THE SAME MODELING PROBLEM MAY HAVE DIFFERENT PERSPECTIVES INTO ITS RESOLUTION AND CAN CERTAINLY COME UP WITH DIFFERENT, VALID ALTERNATIVE SOLUTIONS.

OVERVIEW OF THE MODELING PROCESS



This guide will help you understand each of the components of math modeling. It's important to remember that this isn't necessarily a sequential list of steps; math modeling is an iterative process, and the key steps may be revisited multiple times, as we show in Figure 1.

- **Defining the Problem Statement** Real-world problems can be broad and complex. It's important to refine the conceptual idea into a concise problem statement which will indicate exactly what the output of your model will be.

- **Making Assumptions** Early in your work, it may seem that a problem is too complex to make any progress. That is why it is necessary to make assumptions to help simplify the problem and sharpen the focus. During this process you reduce the number of factors affecting your model, thereby deciding which factors are most important.

- **Defining Variables** What are the primary factors influencing the phenomenon you are trying to understand? Can you list those factors as quantifiable variables with specified units? You may need to distinguish between independent variables, dependent variables, and model parameters. In understanding these ideas better, you will be able both to define model inputs and to create mathematical relationships, which ultimately establish the model itself.

- **Getting a Solution** What can you learn from your model? Does it answer the question you originally asked? Determining a solution may involve pencil-and-paper calculations, evaluating a function, running simulations, or solving an equation, depending on the type of model you developed. It might be helpful to use software or some other computational technology.

- **Analysis and Model Assessment** In the end, one must step back and analyze the results to assess the quality of the model. What are the strengths and weaknesses of the model? Are there certain situations when the model doesn't work? How sensitive is the model if you alter the assumptions or change model parameters values? Is it possible to make (or at least point out) possible improvements?

- **Reporting the Results** Your model might be awesome, but no one will ever know unless you are able to explain how to use or implement it. You may be asked to provide unbiased results or to be an advocate for a particular stakeholder, so pay attention to your point of view. Include your results in a summary or abstract at the beginning of your report.

We will address the components in more detail one by one, but we note again that this should not be thought of as a checklist for modeling. Throughout the process of building your model, you'll likely move back and forth among the components. Take careful notes as you go; it's easy to get caught up in the modeling process and forget what you've done along the way!

PRIMARY EXAMPLES USED THROUGHOUT THIS GUIDE

We demonstrate the modeling process by looking at three modeling questions in detail. We state those problems directly below and then explore them throughout the remainder of this guide.

 **WASTE NOT, WANT NOT: PUTTING RECYCLABLES IN THEIR PLACE**

(A selection from Moody's Mega Math Challenge: 2013 Problem. The full question and a solution paper submitted by Team 1356 from Montgomery Blair High School, Silver Spring, Maryland, coached by David Stein and with student members Alexander Bourzutschky, Alan Du, Tatyana Gubin, Lisha Ruan, and Audrey Shi, is included as Appendix B.)

Plastics are embedded in a myriad of modern-day products, from pens, cell phones, and storage containers to car parts, artificial limbs, and medical instruments; unfortunately, there are long-term costs associated with these advances. Plastics do not biodegrade easily. There is a region of the Northern Pacific Ocean, estimated to be roughly the size of Texas, where plastics collect to form an island and cause serious environmental impact. While this is an international problem, in the U.S. we also worry about plastics that end up in landfills and may stay there for hundreds of years. To gain some perspective on the severity of the problem, the first plastic bottle was introduced in 1975 and now, according to some sources, roughly 50 million plastic water bottles end up in U.S. landfills every day.

Plastics aren't the only problem. So many of the materials we dispose of can be recycled. Develop a mathematical model that a city can use to determine which recycling method it should adopt. You may consider, but are not limited to:

- Providing locations where one can drop off pre-sorted recyclables
- Providing single-stream curbside recycling
- Providing single-stream curbside recycling in addition to having residents pay for each container of garbage collected

Your model should be developed independent of current recycling practices in the city and should include some information about the city of interest and some information about the recycling method. Demonstrate how your model works by applying it to each of the following cities: Fargo, North Dakota; Price, Utah; Wichita, Kansas.



OUTBREAK? EPIDEMIC? PANDEMIC? PANIC?

We all dread getting sick. Years ago, illness didn't spread very quickly because travel was difficult and expensive. Now thousands of people travel via trains and planes across the globe for work and vacation every day. Illnesses that were once confined to small regions of the world can now spread quickly as a result of one infected individual who travels internationally. The National Institutes of Health and the Centers for Disease Control and Prevention are interested in knowing how significant the outbreak of illnesses will be in the coming year in the U.S.



WILL IT THRILL ME?

Amusement parks are typically open during the summer months, when the heat and humidity are almost unbearable. The lines for the most popular rides can sometimes be hours long, leaving you to decide whether you should spend your limited time at the park waiting to ride the newest, most popular roller coaster (with the longest line) or instead riding several, possibly less exciting, roller coasters.

Unfortunately there is no real metric for scoring roller coasters, although an extensive database exists with information about many rides (see *rcdb.com*). Innovative roller coaster engineers certainly set out to design a thrilling roller coaster, but what makes a roller coaster exciting and fun? Create a mathematical model that ranks roller coasters according to a thrill factor that you define.

2. DEFINING THE PROBLEM STATEMENT

Modeling problems are often open ended. Some math modeling problems are clearly defined, while others are ambiguous. This means there is an opportunity for creative problem solving and interpretation. In some cases, it is up to the modeler to define the outputs of the model and which key concepts will be quantified. Defining the problem statement requires some research and brainstorming. The goal is a concise statement that explains what the model will predict.

To see how a math modeling question can be interpreted in different ways, consider the roller coaster problem proposed earlier: rank roller coasters according to how thrilling they are. The word “thrilling” here is open to several interpretations. There are many reasonable possibilities in defining and quantifying “thrilling.”

For example, one student’s definition of a thrilling ride may be a combination of the maximum height and the number of loops, while another student values a combination of length of a ride and the maximum speed. If these individuals ranked the same list of roller coasters, their ranking systems would likely produce different results, neither of which would be “the” correct ranking. The modeler has room to be creative in deciding how to define “thrilling” but must make sure that no matter what definition she decides upon, there is a systematic ranking that incorporates quantifiable (i.e., measurable) aspects of a roller coaster.

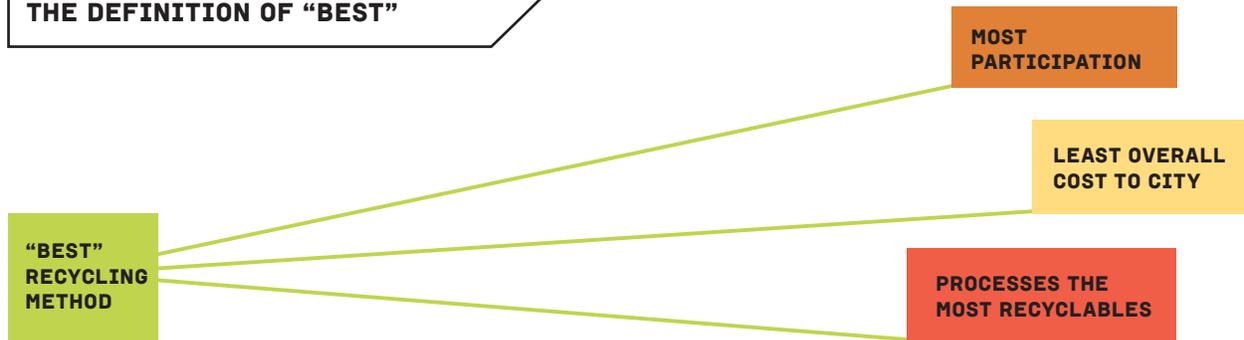
Perhaps you’re thinking that the reason the students above didn’t come up with “the” one correct ranking with either of the previous models is because neither

of those models incorporate sufficiently sophisticated mathematics. Suppose that we can leverage tools from mathematics and physics to help answer this question. Given the design of a particular roller coaster, we might compute, among other things, velocities and g-forces a rider would experience. Even with this information in hand, it’s not obvious how to use that information to rank roller coasters.

Consider four different roller coasters (A, B, C, and D). Coaster A has a larger maximum velocity than B, but B has a higher average velocity. Which is more thrilling? How would these two rank against roller coaster C, which attains a g-force twice as large as A’s or B’s but only does so for 10 seconds of the entire ride? Suppose that roller coaster D never reaches that g-force but sustains g-forces only .5 g less for more than 50 seconds. Which is more thrilling? **The modeler must choose a definition for thrilling.** Eventually, when communicating the results, a modeler will need to explain why decisions were made and will discuss the strengths and weakness of the model.

In the previous discussion we mentioned just a few measurable aspects of roller coasters that one could use to define “thrilling,” including maximum height, the number of peaks, the maximum velocity, or some combination of these. Where does one get a list like this? They come from a process we refer to as brainstorming. Brainstorming is part of the problem-solving process where spontaneous ideas are allowed to flow without evaluation and interruption.

**FIGURE 2
EXAMPLE OF MIND MAP TO EXPLORE
THE DEFINITION OF “BEST”**



The roller coaster example demonstrates that brainstorming at the beginning of a project is an essential process that helps reveal different directions that the math model can take. A brainstorming session may include listing all of the things that make a roller coaster thrilling and then digging deeper to see how those properties are measured. At the beginning of the process, however, one may want to just let the ideas flow and then prune the list later after determining what resources are available. This process is related to making assumptions, which we will talk about in more detail in the next section.

We'll look at the brainstorming process in detail by showing how it might work within the context of the recycling problem. In this problem, we want to determine which recycling method would be best for a city to adopt. The word “best” needs to be clearly defined, and there are multiple ways to do that. Let's imagine that we are on a team that works together to discuss this, and we think of three possible ways to define “best” in this problem.

In order to organize our thoughts, we might use a mind map, as in Figure 2, to give us a visual representation of our initial round of brainstorming. A mind map is a tool to visually outline and organize ideas. Typically a key idea is the center of a mind map and associated ideas are added to create a diagram that shows the flow of ideas. In Figure 2, we focus on the definition of “best,” with three possible definitions

branching off to be further explored. From here, we can focus our attention on one of the three branches at a time. Let's think about the least-cost option first. We probably can't determine how much any recycling program costs without knowing more about the recycling program, so a good place to start is to ask the question “What kinds of recycling programs exist?” If we aren't familiar with different types of recycling, we might need to do some research to see what kinds of programs exist.

If you are working on a long-term modeling project and you have lots of time, you'll want to do an extensive search to find learn everything you can about the problem. You'll also want to find out if others have considered modeling this situation. If you are working on a problem and you have a fairly short time frame, you'll need to be careful to not spend all of your time on the internet researching the problem. Instead, do a quick, preliminary internet search to get a broad perspective (without getting too far into the “weeds”).

Suppose that the list of recycling methods consists of drop-off center, curbside single-stream, curbside (presorted), and pay-as-you-throw. Next, we need to consider the costs. Let's focus on one of the branches, say single-stream curbside pick-up of recyclables. We then ask ourselves, “What contributes to cost for this method?” Then we ask, “For each of those costs, what is the dependence on the properties of the city?”

2: DEFINING THE PROBLEM STATEMENT

FIGURE 3
POSSIBLE MIND MAP UNDER THE ASSUMPTION
THAT “BEST” MEANS LEAST COST

A possible final mind map for the least-cost approach is shown in Figure 3.

Although we will not include the details here, you can imagine that we could proceed in a similar fashion for each of the three definitions of “best.” We would then choose one of the three possibilities, define the problem statement in terms of this choice, and move forward from there to develop a model. During the brainstorming process, explore the problem from different perspectives as if you had access to all the data you could ever need. In the next section, *Making Assumptions*, we’ll discuss exactly what you can do if you can’t find all of the data you need. Don’t discount any idea simply because you don’t think you’ll be able to find sufficient data.

One of the most important aspects of brainstorming is to let the ideas flow freely, especially if done in a group. It is best at this initial phase to stay positive and be open-minded. This part of the modeling process is about creativity, so it is important that there is no criticism of anyone’s ideas or suggestions. What seems like a ridiculous approach may later seem innovative after some more thought, so make note of everything! Also, even if your idea isn’t perfect, it might inspire someone else to come up with an even better suggestion. After you’ve explored the problem and considered several possible approaches, you can step back and look at the possible ways a model might be constructed. Your intuition will help you analyze your brainstorming results and decide on a reasonable problem statement.



2: DEFINING THE PROBLEM STATEMENT

IN SUMMARY

- ① OFTEN MATH MODELING QUESTIONS ARE WORDED IN WAYS THAT ALLOW FOR MULTIPLE APPROACHES, SO YOU SHOULD DEVELOP A CONCISE RESTATEMENT OF THE QUESTION AT HAND.
- ② FOCUS ON SUBJECTIVE WORDS THAT CAN BE INTERPRETED IN DIFFERENT WAYS. ALSO, IDENTIFY WORDS THAT ARE NOT EASILY QUANTIFIED. EXAMPLES INCLUDE BEST, THRILLING, EFFICIENT, ROBUST AND OPTIMAL.
- ③ EXPLORE THE PROBLEM BY DOING A COMBINATION OF RESEARCH AND BRAINSTORMING, KEEPING IN MIND YOUR TIME CONSTRAINTS.
- ④ KEEP AN OPEN MIND AND A POSITIVE ATTITUDE; YOU CAN PRUNE OUT IDEAS LATER THAT ARE NOT REALISTIC.
- ⑤ BRAINSTORMING SHOULD BE APPROACHED AS IF YOU HAD ACCESS TO ANY DATA YOU NEED.
- ⑥ VISUAL DIAGRAMS, SUCH AS MIND MAPS, CAN BE A POWERFUL TOOL LEADING TO THE STRUCTURE OF THE MODEL. CONSIDER USING THE WEBSITE FREEMIND (http://freemind.sourceforge.net/wiki/index.php/Main_Page) [5].
- ⑦ IN THE END, YOU SHOULD HAVE A CONCISE STATEMENT THAT EXPLAINS WHAT THE MODEL WILL MEASURE OR PREDICT.

ACTIVITY

CREATE A MIND MAP FOR THE DISEASE-SPREADING PROBLEM.

3. MAKING ASSUMPTIONS

In presenting any scientific work to others, you need to explain how the results were achieved with explicit details so that they can be repeated. If you are explaining a chemistry experiment, for example, you need to list (among other things) which chemicals were used in what quantities and in what order. Other chemists would expect similar results only when they used the same chemicals and procedure.

The list of assumptions for a mathematical model are as critical as the experimental procedure in performing a chemistry experiment. The assumptions tell the reader under what conditions the model is valid. Making assumptions can be one of the most intimidating parts of the modeling process for a novice, but it need not be! Assumptions are necessary and help you make a seemingly impossible question much more tractable.

Many assumptions will follow quite naturally from the brainstorming process. For the recycling problem, some of our assumptions follow directly from the questions we asked during the brainstorming session, as on the following page.

Let's further examine the assumption about how many people would make use of drop-off centers (termed "likelihood of participation"). The two extremes would be to assume that the 100% of the people near a recycling center would use it or that none would use it. Neither of these seems like a reasonable assumption, so what would be a better assumption? The students whose solution to this problem appears in Appendix B decided

they would do some investigation and see if there has been any successful research on participation rates in drop-off centers. They found a study that had been done in Ohio that estimated about 15% of households participated in drop-off center recycling, and made an assumption that this rate would hold in every city across the U.S.

One might ask if it is safe to assume that across the U.S. 15% of households will participate in drop-off center recycling if it is available. Is it true that residents of Arizona will behave the same way residents of Ohio do? Certainly some cities would garner a participation rate much higher than 15%, while other cities would have a significantly lower participation rate. In fact, what are the chances that any city would actually have a participation rate of exactly 15%? In some sense, one might say that assigning one participation rate to every city across the U.S. is a ridiculous assumption.

In response to that line of thinking, remember two things. First, remember that one must make assumptions in order to make a model. It is not practical or feasible to poll every citizen of every city to determine who will bring recyclables to a drop-off center. If we had to rely on data with that level of certainty at every juncture of the modeling process, we would never get any work done. It's practical and important to make reasonable assumptions when we cannot find data.

3: MAKING ASSUMPTIONS



Second, you are developing a model that is intended to help one understand some complex behavior or assist in making a complex decision. It is not likely to predict the exact outcome of a situation, only to help provide insight and predict likely outcomes. When you provide a list of your assumptions, you've done your part to inform anyone who might use your model. They can decide whether they think your assumption is or is not appropriate to model the behavior they are interested in predicting. In the *Analysis and Model Assessment* section, we'll discuss in more detail ways in which you can examine some of the impacts of your assumptions.

It's entirely possible that you may search and search and never find the data you need to make an "educated" guess about a parameter in your model. That's fine; simply make a note in your write-up that future work might include further investigation in that area. If Team 1356 had not found any estimates for recycling rates, they might have assumed that the recycling rate was 50% in the absence of other data (since it's the mean of the two extreme cases). That would have been a better assumption than either of the extremes (all residents recycle or no residents recycle). They also might have determined that 25% seemed reasonable (based on their own experiences or intuition) and moved forward with that number. All of these are appropriate as long as they are included as assumptions.

Choice of assumptions may also be dictated by the mathematical tools available. Both the National Institute of Health and the Centers for Disease Control and Prevention use mathematical modeling to help them understand the spread of infectious diseases. While their models may be quite sophisticated, they are actually built upon many of the simple principles we will discuss here, which evolve from relatively few assumptions. Let's focus on determining the number of people who have the disease over time by considering models at multiple math levels.

One of the simpler models for disease propagation can be created if we assume that the disease spreads at a constant rate. For example, we might assume that each person who has the disease will spread the disease to 3 people per day or that each person spreads the disease to just 1 person every 5 days. As we move forward, we will refer to this as the *constant-rate disease model*.

Transmission rate drives the spread of disease, and the assumption that it remains constant over time seems unlikely for the duration of the disease. If we have knowledge of calculus and differential equations, we can arrive at another model that accounts for varying transmission rate.

3: MAKING ASSUMPTIONS

In order to decide how the transmission rate should change with time, it might be helpful to think about the mechanism behind disease transmission: infected people somehow in contact with susceptible individuals. It makes sense, then, to believe that the transmission rate depends on how many people are infected and how many people are susceptible. We might assume that the transmission rate is directly proportional to the product of the number of people infected and the number of people who are susceptible. We will refer to this model as the *varying-rate disease model*. We will revisit both disease models in later sections.

Some assumptions are made at the beginning of the modeling process, while others are made as you proceed through the modeling process. The modeling process is iterative;

it is legitimate to make a reasonable assumption, determine how it affects the model moving forward, and make adjustments to improve the outcome. You can see an example of this later in the *Analysis and Model Assessment* section. Make a careful list all of the

assumptions you make along the way; a good modeling paper includes a list of assumptions in the write-up. Additionally, keep track of all the resources used so that you can create a bibliography.

With all of these options, how does one know which is the best assumption to make? There is no easy answer to this question; the most important thing is to acknowledge the assumptions you've made and, when appropriate, discuss the limitations

that might arise from your assumptions.

HOW DOES ONE KNOW WHICH IS THE "BEST" ASSUMPTION TO MAKE? THERE IS NO EASY ANSWER TO THIS QUESTION; BUT BE SURE TO ACKNOWLEDGE THE ASSUMPTIONS YOU'VE MADE AND DISCUSS THEIR LIMITATIONS.

IN SUMMARY

- 1 ASSUMPTIONS OFTEN COME NATURALLY FROM THE PROCESS OF BRAINSTORMING AND DEFINING THE PROBLEM STATEMENT.
- 2 YOU SHOULD DO SOME PRELIMINARY RESEARCH AND MAY FIND DATA TO HELP YOU MAKE ASSUMPTIONS. IN THE ABSENCE OF RELEVANT DATA, MAKE A REASONABLE ASSUMPTION AND JUSTIFY THE ASSUMPTION IN YOUR WRITE-UP.
- 3 DIFFERENT ASSUMPTIONS CAN LEAD TO DIFFERENT, EQUALLY VALID MODELS AT DIFFERENT MATHEMATICAL LEVELS.
- 4 NOT ALL ASSUMPTIONS ARE MADE DURING THE INITIAL BRAINSTORMING. SOME COME AS YOU MOVE THROUGH THE MODELING PROCESS. KEEP TRACK OF THE ASSUMPTIONS YOU MAKE AND INCLUDE A LIST OF ASSUMPTIONS IN YOUR WRITE-UP OF THE MODEL.

ACTIVITY

BUILD ON THE BRAINSTORMING FROM THE PREVIOUS SECTION ABOUT THE ROLLER COASTER MODEL. CERTAINLY WE DID NOT UNCOVER ALL THE WAYS IN WHICH ROLLER COASTERS ARE CONSIDERED THRILLING. DEFINE THE PROBLEM STATEMENT IN YOUR OWN WORDS, BASED ON YOUR UNDERSTANDING OF THE PROBLEM. FINALLY, TAKE YOUR WORK ONE STEP FURTHER AND LIST THE ASSUMPTIONS ON WHICH YOU COULD BUILD YOUR MODEL.

4. DEFINING VARIABLES

With the problem statement clearly defined and an initial set of assumptions made (a list that will likely get longer), you are ready to start to define the details of your model. Now is the time to pause to ask what is important that you can measure. Identifying these notions as variables, with units and some sense of their range, is key to building the model.

The purpose of a model is to predict or quantify something of interest. We refer to these predictions

as the outputs of the model. Another term we use for *outputs* is *dependent variables*. We will also have *independent variables*, or *inputs* to the model. Some quantities in a model might be held constant, in which case they are referred to as *model parameters*. Let's look at a few simple examples that will help you distinguish between these concepts. We'll also see how they depend on your viewpoint and the problem statement.

ANOTHER TERM FOR OUTPUT.

THE INPUTS OF THE MODEL.

The purpose of a model is to predict or quantify something of interest. We refer to these as the outputs of the model.

EXAMPLE 1: PAINTING A HOUSE

Suppose that we plan to paint a house. We'll need to know the dimensions of the house so that we can find the surface area, SA (ft^2). We also need to know the efficiency, E (ft^2/gal), of the paint, which tells us how many square feet a gallon of the paint can cover. Keep in mind that the efficiency varies from brand to brand. We let V (gal) be the volume of paint in gallons that we need. Here, knowing the units for efficiency can help reveal the relationship between the variables:

$$E = \frac{\text{surface area}}{\text{volume}} = \frac{SA}{V}$$

NOTICE THAT WE CAN REWRITE THIS RELATIONSHIP AND USE ANY OF THE FOLLOWING THREE EQUIVALENT RELATIONSHIPS:

- ① $E = SA / V$
- ② $SA = E \cdot V$
- ③ $V = SA / E$

Whether something is a dependent or independent variable or a parameter often relies on the perspective of the modeler and the problem statement. Imagine that you own a painting company and always use CoversItAll brand paint. When a client hires you, you take measurements of the house, and then you want to know how much paint you'll need to complete the job. In this scenario, the efficiency of CoversItAll paint is constant, so that is a model parameter. You would use equation (3), with the surface area of the home as the input and the volume of paint needed as an output. Therefore, SA is the independent variable, V is the

dependent variable, and E is a constant model parameter.

Suppose instead that you are a homeowner and

WHETHER SOMETHING IS A DEPENDENT OR INDEPENDENT VARIABLE OR A PARAMETER OFTEN COMES FROM THE PERSPECTIVE OF THE MODELER AND THE PROBLEM STATEMENT.

want to choose from among five brands of paint to buy to minimize the amount necessary to paint your house. Under this scenario, the surface area of your house is a constant, so this is treated like a model parameter. If we know the efficiency of each of the five brands of paint, we would again use equation (3), but this time with E as the input variable and volume as the output. Thus,

in this case, E is the independent variable and V is the dependent variable.

EXAMPLE 2: CONSTANT-RATE DISEASE MODEL

For another example, recall the constant-rate disease model discussed in the previous section, wherein we want to determine the number of infected individuals at any given time. Using this problem statement, let t denote the time in days. This will be the input (independent variable). Let $I(t)$, the number of infected individuals at time t , be the output of the model (the dependent variable). In this constant-rate disease model, we assume that each person who has the disease will spread the disease to a certain number of people in a given fixed time period. We define the parameter τ to be the time period during which each infected person will transmit the disease to r other people (so r is also a parameter). Further, we might define a parameter I_0 to be the initial number of infected individuals. In other words, $I(0) = I_0$. For example, if we have $I_0 = 10$, $\tau = 2$ days, and $r = 3$, then we are considering a population starting with 10 infected individuals, where each infected person transmits the disease to 3 uninfected individuals every 2 days.

These simple examples show that the problem statement will guide what the dependent variable (i.e., your model output) is going to be. Dependent variables and parameters will often be determined by both your assumptions and the availability of information. The key idea is that independent variables cause a change in the dependent variables. Let's look at a more complex model that demonstrates how submodels may be needed to provide input to the overarching model.

The key idea is that independent variables cause a change in the dependent variables.

EXAMPLE 3: DETERMINING RECYCLING COSTS

For the recycling problem, we seek a model that will predict the cost for a city to implement and run a recycling program. Hence, the output of the model should be in dollars. What will the inputs to the model be? We can use the results of our brainstorming session to help us. Let's consider the model to determine the cost of a drop-off center. Based on the brainstorming shown in Figure 3, the cost of using drop-off centers in a given city could depend on how many are needed, the operational cost, the likelihood that people will participate, and the possibility of refunds or incentives.

Let's consider the approach taken in the solution provided in this guide. These students calculated the cost of a drop-off center, based on the cost to maintain the center, as well as the revenue the center would create. This latter fact depends directly on the participation rate by the local population. So, the probability of participation by a single household was needed first. Determining the single-household probability rate is an example of how a submodel can be used to generate input to the main model.

Continuing this line of thought, the students based the likelihood that a household would participate on its distance from a drop-off center. Thus, the distribution of houses throughout the city and the location of drop-off centers must be understood. This team assumed that each city was a square and that houses were aligned on grids. To determine the placement of the drop-off centers within the city grid (which will actually determine how many will be needed), students calculated a maximum distance, d , that citizens would

be willing to drive per week to a drop-off center. Using that distance, drop-off centers were placed on the grid so that they did not overlap yet the entire city would be covered.

Thus, for this approach, note that d is an input to determine where to place the drop off-centers, yet d itself needs to be determined first (because it certainly isn't clear what a reasonable value of d might be nor is there one known value that suits everyone in the United States). Therefore, ultimately we can use a math modeling approach to determine d based on some assumptions as well. This team decided that d depends on the cost to get to a drop off-center, and this would depend on the price of gas, gas mileage for a typical car, and the amount that a household would be willing to pay to recycle per month. Values for these model parameters were found in the literature through some digging into available resources.

To summarize their approach, they assumed that the following were model parameters to find d :

- People would be willing to pay \$2.29 to recycle per month or \$0.53 per week.
- People would make biweekly trips to the center.
- The average price of a gallon of gas is \$3.784.
- The average mileage of a passenger car is 23.8 miles/gallon.

4: DEFINING VARIABLES

$$d = \left[\frac{(\$0.53/\text{week}) \cdot (23 \text{ miles/gallon})}{\$3.784/\text{gallon}} \right] / 2 = 1.66 \text{ miles/week.}$$

With a value of d in place, students were able to take into consideration the size of the city grid and then partition the city to place the appropriate number of recycling drop-off centers to cover the entire city. Next, the team wanted to propose a method for predicting the likelihood of participation, but this couldn't be done until the previous submodels were developed.

Note that the inputs, or independent variables, for their cost model for drop-off centers were the area of the city, the population, the average number of people in a household, the maximum distance citizens would be willing to drive, and the number of drop-off centers needed. However, more modeling was needed to determine values for many of these inputs, as we demonstrated with the input, d . Notice that these are specific details that are implied from the brainstorming shown in the mind map in Figure 2 but at this phase in the process, more detail (and additional brainstorming and assumptions) is needed.

IN SUMMARY

- 1 THE PROBLEM STATEMENT SHOULD DETERMINE THE OUTPUT OF THE MODEL. THE OUTPUT VARIABLES THEMSELVES WILL BE DEPENDENT VARIABLES.
- 2 THE RESULTS OF THE INITIAL BRAINSTORMING CAN PROVIDE INSIGHT INTO WHICH VARIABLES SHOULD BE INDEPENDENT VARIABLES AND WHICH SHOULD BE FIXED MODEL PARAMETERS.
- 3 KEEP TRACK OF UNITS BECAUSE THEY CAN REVEAL RELATIONSHIPS BETWEEN VARIABLES.
- 4 YOU WILL LIKELY NEED TO DO SOME RESEARCH AND MAKE ADDITIONAL ASSUMPTIONS TO OBTAIN VALUES FOR NECESSARY MODEL PARAMETERS.
- 5 SUBMODELS OR MULTIPLE MODELS MAY BE NEEDED TO GENERATE SOME OF THE MODEL INPUT.

ACTIVITY

DETERMINE THE DEPENDENT AND INDEPENDENT VARIABLES FOR RANKING ROLLER COASTERS BASED ON HOW THRILLING THEY ARE. WHAT ARE SOME POSSIBLE MODEL PARAMETERS?

5. BUILDING SOLUTIONS

Now that you have an initial mathematical model, you will need to use that model to generate preliminary answers to the question at hand. The approach you take of course depends on the type of model you have and your background in mathematics. It may involve simply considering some different values of certain parameters to see how the output changes, it may involve techniques from calculus or differential equations, or it may involve using graphs to understand trends in data. In this chapter we will give you some strategies for choosing how to solve your problem.

When you first approach any mathematical problem, you often look into your personal tool kit for a mathematical technique to use. Sometimes, if we start with the incorrect approach, a better approach will naturally emerge. So, the important thing is to just tackle it and see what happens! The following questions may help you.

- **HAVE I SEEN THIS TYPE OF PROBLEM BEFORE?**
- **IF SO, HOW DID I SOLVE IT? IF NOT, HOW IS THIS PROBLEM DIFFERENT?**
- **DO I HAVE A SINGLE UNKNOWN, OR IS THIS A MULTI-VARIABLE PROBLEM WITH MANY INTERDEPENDENT VARIABLES?**
- **IS THE MODEL LINEAR OR NONLINEAR?**
- **AM I SOLVING A SYSTEM OF EQUATIONS SIMULTANEOUSLY, OR CAN I SOLVE SEQUENTIALLY?**
- **WHAT SOFTWARE OR COMPUTATIONAL TOOLS ARE AVAILABLE TO ME?**
- **WOULD A GRAPH OR OTHER VISUAL SCHEMATIC HELP PROVIDE INSIGHT?**
- **COULD I APPROXIMATE MY COMPLICATED MODEL WITH A SIMPLER ONE?**
- **CAN I HOLD SOME VALUES CONSTANT AND ALLOW OTHERS TO VARY TO SEE WHAT IS GOING ON?**

5: BUILDING SOLUTIONS

Certainly there are times when it is clear what mathematical technique is required (for example, factoring, finding zeros of a polynomial or a function, integrating a function, simulating a model over time to understand how output evolves, etc.). Other times, when it is not clear how to proceed, it may be helpful to analyze simple examples, special cases, or related problems. Even a “guess-and-check” approach can sometimes provide some deep insight. Mathematical experiments may be facilitated with a graphing calculator or computational software such as Excel, Mathematica, or Maple.

Multiple approaches can be taken to build a solution. We will show several approaches for each of the following examples, which appear in order of increasing mathematical level. We also show how to leverage software to assist you in finding solutions.

EXAMPLE 1: ACID RAIN

Let’s consider approaching a general problem by building different models that may lead to different results. Suppose we want to determine how acid rain is impacting the water resources in your town. We already know from our previous section that this open-ended statement needs to be refined into a concise problem statement. Let’s suppose we developed the following problem statement after some brainstorming and mind-mapping: *Measure the levels of sulfur dioxide (SO_2) and nitrogen monoxide (NO) at four different locations and use the results to determine the best place for water to be extracted.*

APPROACH 1.

Ranking Given the measurements for each location, rank the locations as safest, second safest, third safest, and least safe. This allows a qualitative approach but incorporates at least some quantitative analysis, since it requires us to assess the importance of SO_2 vs. NO and define the term *safe*.

APPROACH 2.

Equation-Based Solution We can assess the difference in importance between SO_2 and NO , and create an equation that assigns a score to each location. The location with the highest score wins. This requires algebraic modeling and manipulation as well as ideas of proportionality.

APPROACH 3.

Qualitative Comparison We can decide if any of the sites is too polluted. For example, if one site has the highest levels of both SO_2 and NO , then that site should not be used. This reduces the problem to choosing among three sites, and the same ideas can be applied to reduce them to two and then to a final one.

EXAMPLE 2: CONSTANT-RATE DISEASE MODEL

Often, when modeling real-world phenomena, we are interested in forecasting future values. That is, we want to explore how the value of something we care about changes over time. We will discuss several approaches that we could take to find a solution to the constant-rate disease model (in which we want to determine the number of people who are infected with a disease under the assumption that each person afflicted with a disease spreads it to exactly r people over a given time period τ). For this example, we'll let $I_0 = 10$, $\tau = 2$ days, and $r = 5$. In other words, we have a situation in which 10 people were infected with a disease that they, and all other future individuals who become infected, will each transmit to exactly 5 susceptible individuals every 2 days.

APPROACH 1.

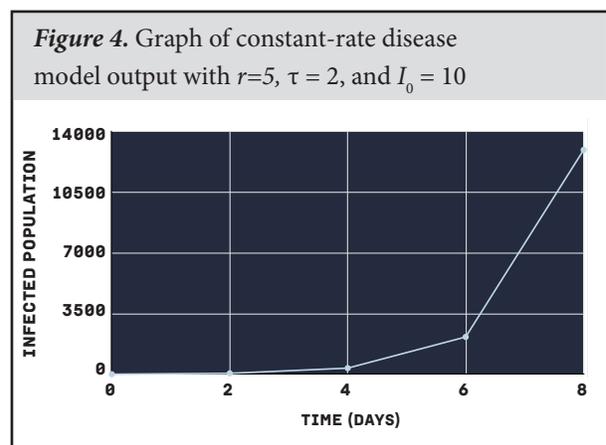
Computation by Hand We start by doing some simple computations by hand to determine whether a pattern emerges. After 2 days, we have the original 10 people who have the disease, but we also have 50 more, because each of the 10 infected individuals has spread the disease to 5 others. Hence, when $t = 2$, $I = 60$. After 2 more days, we have $I = 60 + (5 \times 60) = 360$, and so on. We could organize our numbers by putting them in a table, as in Table 1.

<i>Table 1.</i> Constant-rate disease model computations, by hand	
t , In Days	I , Infected Population
0	10
2	60
4	360
6	2160
8	12960
10	77760
12	466560

APPROACH 2.

Computation via Technology Straightforward iterative calculations such as those needed for this problem are easily “programmable” in spreadsheet software such as Microsoft Excel. Excel is a useful tool for visualizing step-by-step discretization, and it can allow you to see the value of each variable at each time step in table form. If you don't know how to use Excel, you might find resources or videos by doing an internet search on performing iterative calculations in Excel.

If we've done our computations in Excel, we can also create a plot of our solution easily, as in Figure 4.



EXAMPLE 2: CONSTANT-RATE DISEASE MODEL (CON'T.)

APPROACH 3.

Pattern identification We might notice as we perform computations by hand or by using technology that if we know the number of infected individuals at some time step, then we can get the infected population at the next time step by multiplying by 6. Let's see why this happens.

At time $t = 0$, we have 10 infected individuals. At time $t = \tau = 2$ days, we have the original 10 infected individuals plus 50 more, for a total of 60.

For this set of parameters: *For a generic set of parameters:*

$$\begin{aligned} I(\tau) &= 10 + 5 \cdot 10 & I(\tau) &= I_0 + \tau I_0 \\ &= (1 + 5) 10 & &= (1 + r) \cdot I_0 \\ &= 6 \cdot 10. \end{aligned}$$

Continuing, at time $t = 2\tau = 4$ days, we have the 60 from the previous time step, plus 5 times 60 more.

For this set of parameters: *For a generic set of parameters:*

$$\begin{aligned} I(2\tau) &= 60 + 5 \cdot 60 & I(2\tau) &= I(\tau) + r \cdot I(\tau) \\ &= (1 + 5) 60 & &= (1 + r) \cdot I(\tau) \\ &= 6 \cdot 60. \end{aligned}$$

What does the formula look like for $I(3\tau)$? (Try it!)

Now we see where multiplication by 6 comes from, but we might be able to see an even deeper formula emerge. We can actually substitute our expression for $I(\tau)$ into the equation the equation for $I(2\tau)$, as below.

For this set of parameters: *For a generic set of parameters:*

$$\begin{aligned} I(2\tau) &= 6 \cdot 60 & I(2\tau) &= (1 + r) I(\tau) \\ &= 6 \cdot (6 \cdot 10) & &= (1 + r) \cdot (1 + r) I_0 \\ &= 6^2 \cdot 10. & &= (1 + r)^2 I_0. \end{aligned}$$

You should verify that for $I(3\tau)$ we have the following equations.

For this set of parameters: *For a generic set of parameters:*

$$I(3\tau) = 6^3 \cdot 10. \qquad I(2\tau) = (1 + r)^3 I_0.$$

You might see a pattern emerging that would lead you to find a closed form solution for the infected population after n days have passed.

For this set of parameters: *For a generic set of parameters:*

$$I(n\tau) = 6^n \cdot 10. \qquad I(n\tau) = (1 + r)^n I_0.$$

This result, an exponential model, is consistent with both the values in Table 1 and the corresponding graph in Figure 4.

The results of each of the three approaches are perfectly valid model for our stated assumptions, but we now see that the model may be limited in its ability to accurately describe some of the real-world characteristics of disease propagation. In the next section we discuss these questions, and we'll revisit this model as a part of model assessment.

EXAMPLE 3: VARYING-RATE DISEASE MODEL

We haven't defined the variables yet in the varying-rate disease model, so let's do that now and set up the differential equation which is to be solved.

We define the total population to be P , and each member of the population must belong to exactly one of two classes: susceptible, S or infected, I . Hence, $P = S + I$. We assume that the total population remains constant, but the values of S and I change over time, so we might choose to write $P = S(t) + I(t)$.

Recall that in the varying-rate disease model, we consider a population in which disease transmission rate is directly proportional to the product of the number of people infected and the number of people who are susceptible. We can think of transmission rate as the rate at which people become infected, or the rate of change of population $I(t)$. Readers familiar with calculus may recognize this as the derivative. We will denote this rate of change $I(t)$ with respect to time as $\frac{dI}{dt}$. Then we can translate the assumption "disease transmission rate is directly proportional to the product of the number of people infected and the number of people who are susceptible" into the equation

$$\frac{dI}{dt} = kI(t)S(t), \quad (4)$$

where k is a (positive) proportionality constant. The larger the value for k , the larger $\frac{dI}{dt}$ is. So a large k -value indicates a highly contagious disease.

In order to find a solution to the differential equation, it will be helpful if we can get down to only one dependent variable. Since we assumed that the population, P , is constant, then we can take advantage of the relationship $P = I(t) + S(t)$ to write S as $S(t) = P - I(t)$. Then we can rewrite equation (4) as follows:

$$\frac{dI}{dt} = kI(t)(P - I(t)). \quad (5)$$

APPROACH 1.

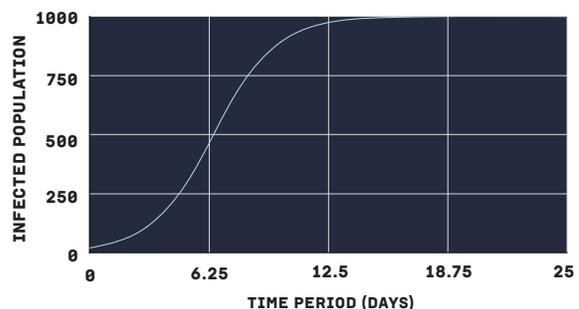
Analytic Solution If you've studied differential equations before, you might recognize that this particular differential equation can be solved analytically using a technique called separation of variables. (See any standard calculus text.) Using this technique and assuming that the initial condition is $I(0) = I_0$, you can find the solution to be

$$I(t) = \frac{PI_0}{I_0 + (P - I_0)e^{-Pkt}}. \quad (6)$$

Alternately, if you have access a symbolic computation tool such as Maple, you can use technology to generate this analytic solution.

With the analytic solution in hand, we can demonstrate how the model behaves by choosing some parameter values to generate a plot. (See Figure 5.) Notice that this model exhibits the same sort of exponential growth in the initial stages of disease propagation as the constant-rate disease model, but the rate slows when there are fewer people left who have not yet contracted the illness. We see that I approaches 1000 as time increases. In generating the plot, we used total population $P = 1000$ (a parameter), so our model predicts that over time the entire population contracts the disease.

Figure 5. Analytic solution of the varying-rate disease model output with $k = 0.0006$, $P = 1000$, and $I_0 = 20$



EXAMPLE 3: VARYING-RATE DISEASE MODEL (CON'T.)

APPROACH 2.

Approximate Solution While we can find an analytic solution for the differential equation above, many differential equations, as well as other equations and systems of equations that describe real-world phenomena, cannot be solved directly. In these situations, it is common to approximate a solution, often referred to as using a “numerical method.” Numerical methods are a powerful tool in modeling, especially if you have not yet had formal training in techniques from calculus and differential equations.

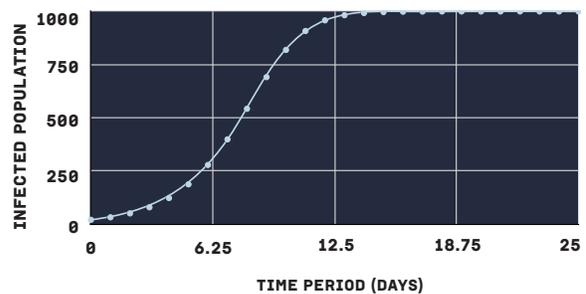
As an example, consider equation (5). Let’s suppose that we did not have the analytic solution (equation (6)) to evaluate at any time we choose. One numerical method would be to try to calculate approximate values of the infected population at specified later times. We know the initial population, $I(0) = I_0$, and we have a model describing the rate of change of the population. Intuitively we should be able to predict the number of infected people at a later time, say $t = 1$. To do this, we can express the change in the infected population over that time frame time as

$$\frac{dI}{dt} \approx \frac{I(1) - I(0)}{1 - 0}$$

and plug what we know into the right-hand side of equation (5). We can then solve for $I(1)$ in terms of I_0 . We should be careful, however. What we really have is an approximation to $I(1)$ because we approximated $\frac{dI}{dt}$, but our goal was to determine approximate values of the infected population, and that is what we’ve accomplished. We could then use our value of $I(1)$ in the same

way to predict $I(2)$, and so on. This numerical method, called the forward Euler method, is easily implemented in Excel.

Figure 6. Numerical solution of the varying-rate disease model output with $k = 0.0006$, $P = 1000$, $I_0 = 20$, and time step $\Delta t = 1$ day



We show a plot of the approximated values of the infected population at different points in time in Figure 6, and you can see the details of how we obtained these results in Appendix A.

Notice that the plot of the numerical solution looks much like the plot of the analytic solution we saw in Figure 5. While we are pleased that the numerical solution does a good job of approximating the analytic solution, we should note that the graphs are not identical. Every numerical method introduces some error, and the forward Euler method is no exception. The study of error is a complicated issue, and we will not attempt to address it here.

SUMMARY OF BUILDING SOLUTIONS

Finding a solution to your model may be achieved by various means, depending on your background knowledge in mathematics and software. Solving by ranking is an excellent method for students who do not have the mathematical training to produce algebraic formulas. Even with an equation, however, there are often multiple ways to arrive at a final answer. Software tools such as Excel can facilitate obtaining solutions. If you do use a numerical approximation technique, it is important to be aware of error that might be introduced. If nothing else seems to help, try guess-and-check.

IN SUMMARY

- 1 HOW YOU BUILD A SOLUTION MAY DEPEND ON WHAT MATHEMATICAL TOOLS ARE AVAILABLE TO YOU.
- 2 THERE IS OFTEN MORE THAN ONE WAY TO TACKLE A PROBLEM, SO JUST START AND SEE WHAT HAPPENS.
- 3 IF YOU DON'T IMMEDIATELY KNOW HOW TO SOLVE THE PROBLEM AT HAND, ASK YOURSELF THE PROVIDED SET OF QUESTIONS TO HELP YOU GET STARTED.
- 4 DIFFERENT SOLUTION METHODS CAN LEAD TO SOLUTIONS OF DIFFERENT NATURES. THIS IS PERFECTLY ACCEPTABLE.

ACTIVITY

CONTINUE WORKING WITH YOUR MODEL FOR A RANKING SYSTEM FOR ROLLER COASTERS DEPENDING ON YOUR DEFINITION OF "THRILLING." USE YOUR RANKING SYSTEM TO RANK AT LEAST 10 ROLLER COASTERS. YOU MAY FIND DATA YOU NEED AT RCDB.COM.

6. ANALYSIS AND MODEL ASSESSMENT

We separate this section into two subsections. The first subsection, *Does My Answer Make Sense?*, gives some quick checks to determine whether your solution is at all reasonable. The second subsection, *Model Assessment*, gives more in-depth techniques to analyze the model.

Often we are so excited that we built and solved a mathematical model that we forget to step back and carefully examine the results. While this is understandable since it took hard work to get to that point, it is essential to ask yourself, “Does my answer make sense?” Sometimes, the results may indicate a mistake in the calculations. Other times you may find that additional or alternate assumptions are needed for the solution to be realistic. If the results do make sense,

then further analysis is needed to assess the quality of the model. Recall that open-ended questions may have more than one solution and that the results depend on the assumptions made and the level of sophistication of the mathematics used. An honest evaluation is necessary to explain when the model is applicable and when it is not. In this section we will talk about ways to analyze your results and how to assess the quality of your model.

DOES MY ANSWER MAKE SENSE?

During the modeling process, you may gather some intuition about what the output will be like. Here we provide some pointers on how to answer the question “Does my answer make sense?” by analyzing the output of the model.

- **Is the sign of the answer correct?** For example, if your disease model is supposed to calculate the number of infected people at a certain time, then clearly an answer of -1000 would not make sense. Carefully check your calculations, especially if you are using software. For example, in Excel it is easy to select the wrong cell when defining a formula. Your model may be correct, but its implementation may be at fault.

- **Is the magnitude of the answer reasonable?** If you are trying to estimate the speed of a car, for example, then it wouldn't make sense if your model predicts a value of 1000 miles per hour. Sometimes, when the magnitude of a number is off, incorrect units may have been used somewhere in the process.

- **Does the model behave as expected?** If the output of your model is visualized with a graph or plot of any kind, then carefully look at the intercepts, the maximum or minimum values, or the long-term behavior. Were you expecting a horizontal asymptote, yet your graph just increases without bound? If you have a data set and believe there is a relationship between two variables, plot the data. A mathematical error in the sign of the slope will be immediately obvious. It could be that you had some assumptions which were neglected, erroneous units, unrealistic parameters, or that the software was used incorrectly.

- **Can you validate the model?** Sometimes it is possible to validate your model using available data. For example, if you used your roller coaster ranking model on the Top Thrill Dragster of Cedar Point, which held the record for the tallest roller coaster in the world and goes 120 mph, and the output said it was only mediocre, then likely your model is not doing what you want it to.

MODEL ASSESSMENT

Now that you have verified that your model is correct, it is time to step back and consider the validity of your model. This includes identifying the strengths and weaknesses of your model and understanding at a deeper level the behavior of the model. Performing a sensitivity analysis, wherein you analyze how changes in the input and parameters impact the output, can contribute to understanding the behavior of your model.

IDENTIFYING STRENGTHS AND WEAKNESSES

Once you are convinced that the output is correct and the model is achieving what you want, assess the quality of model. This assessment needs be included in the write-up about your model to help people understand the conditions under which your model is applicable, which is strongly linked to the assumptions that were made along the way. It is necessary for you to provide an honest, exact assessment of the capabilities of your model.

This is also a chance to highlight the strengths of the model. For example, even if a model was formulated using simple physics, it might require very little expert knowledge in order to provide meaningful insight. This can be a huge advantage over a more complex model that requires the user to program and run

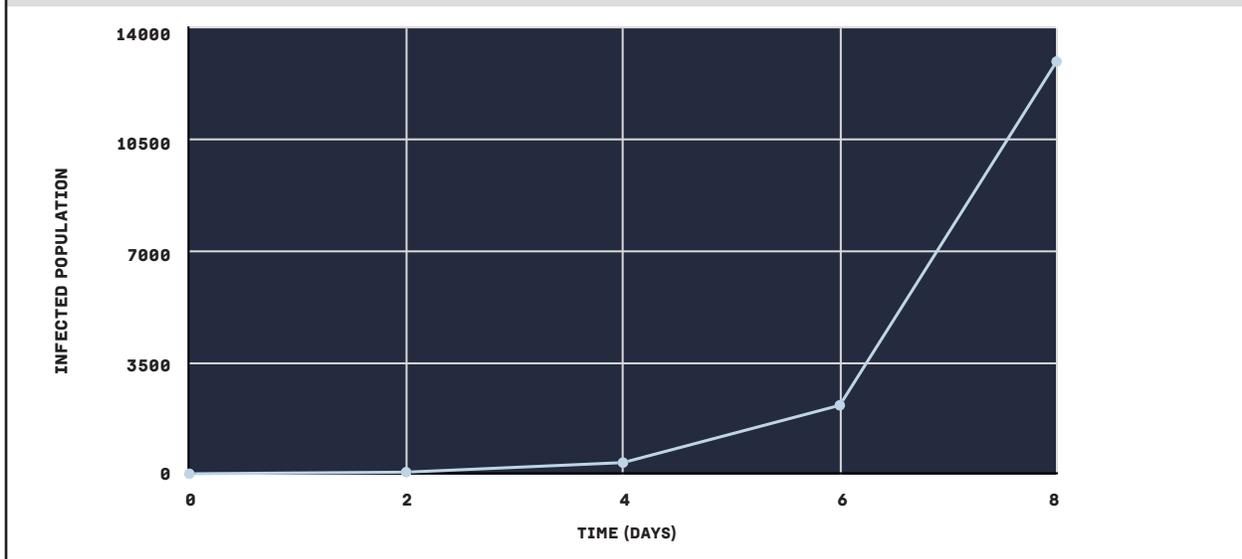
software, research other model parameters in order to fit the model to his or her own needs, or sort through complicated output to be able to draw a conclusion. It is also a strength if the people who might use your model can understand it and have faith in it.

Let's examine this process by looking at the constant-rate disease model. Recall our assumption that each infected person transmits the disease to r people every τ days. From this, we found the following exponential function to describe the infected population after $n\tau$ days:

$$I(n\tau) = (1 + r)^n I_0.$$

Figure 7 shows the graph of the model output for $I_0 = 20$ infected people, $\tau = 2$ days, and $r = 5$.

Figure 7. Graph of constant-rate disease model output with $r = 5$, $\tau = 2$, and $I_0 = 20$



THIS MODEL HAS SOME VALUABLE STRENGTHS:

- This model is easy to describe to others, which means that they might have increased confidence in its output. Allowing full understanding of all parts of the model can be valuable when trying to understand the significance of output given different input values.
- We found an analytic solution. If we have values for I_0 , τ , and r , the function $I(n\tau) = (1 + r)nI_0$ can be used to solve for the number of infected individuals at any time.
- Our model is consistent with our assumptions. Our primary assumption is that the disease is spread at a constant rate. Our model accurately describes this behavior, and therefore we have developed a meaningful solution.

THIS MODEL ALSO HAS A FEW NOTABLE WEAKNESSES:

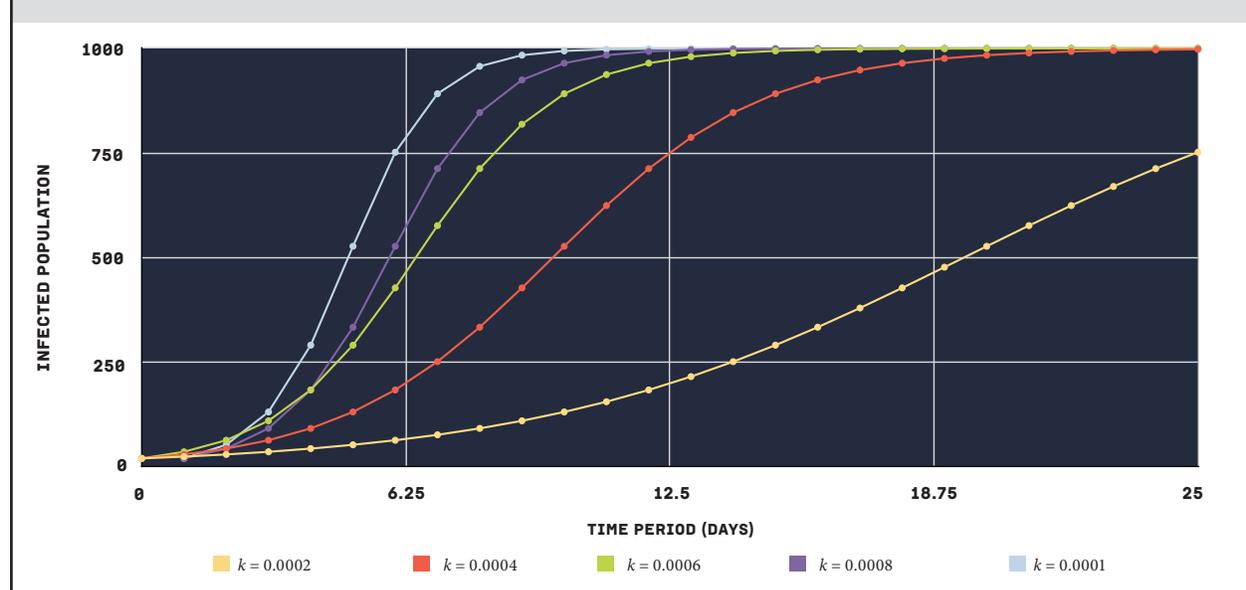
- Our model is quite simple. In addition to being a strength, it is also a weakness. We may be concerned with the constant rate assumption, especially since it is not consistent with our intuition regarding the spread of disease.

- This model is not valid for all types of disease. Our model only has two categories for individuals: susceptible and infected. In real life, after being infected with some diseases a person recovers and then acquires immunity to that specific disease. This model wouldn't accurately describe that situation.
- This model predicts the disease spreads to everyone. In other words, under this set of assumptions, the disease will propagate through a population until every individual has the disease. This (hopefully) is not a reasonable scenario.

In this case, we have a model based on sound mathematics that provides us with good insight into disease propagation but also leads to some questionable real-world outcomes if taken as the final answer. In general, awareness of your model's capabilities leads to better overall solutions. In such awareness, you know when it is appropriate to use your model, and you also have a starting point from which to build future (more specific and/or realistic) models. Identifying your model's weaknesses does not detract from the hard work you've done; it is always preferable for the modeler to acknowledge weaknesses in the model. If the reader identifies weaknesses that the modeler has missed, then the reader's assessment of the model and the modeler suffers.

SENSITIVITY ANALYSIS

Figure 8. Analytic solution of the varying-rate disease model output with $P = 1000$, $I_0 = 20$, and k ranging from 0.0002 to 0.001 in increments of 0.0002



When doing model assessment it is also vital to consider the model's sensitivity to changes in the assumptions and parameters used to build it. A model is considered sensitive with respect to a parameter if small changes in that parameter lead to significant changes in the output. There are several ways to conduct a sensitivity analysis. A simple approach would be to consider a range of values for a certain parameter while keeping all other parameters fixed, calculate the output, and then determine the impact on the output. For example let's take a look at the effect of changes to the transmission rate in the variable-rate disease model with $P=1000$ and $I_0=20$.

Looking at Figure 8, we see how incremental changes to the transmission rate affect the output of our model. In particular, we note that when $0.0006 \leq k \leq 0.001$, the disease has infected most (if not all) of the population after 15 days, but it takes just over 20 days for the disease to reach most of the population when $k = 0.0004$. When $k = 0.0002$, a good proportion of the population still has not been infected after 15 days. What

does this mean? Well, that depends on the problem you are solving. In this example, we learned that changes to k can increase (or decrease) the spread of the disease to the entire population. If a population were infected with such a disease, then this model could demonstrate that a drug that may inhibit the disease transmission rate could create the time needed to develop an effective course of treatment.

In our sensitivity analysis, the changes in k were fairly small (0.0002). How do you know how big (or small) your variation should be? In some cases you may have real-world data to help you make a decision. If that's not available, use common sense as your guide and play around with the numbers to develop intuition. For example, in the variable-rate disease model, we could also hold k constant and investigate whether changes to the initial infected population I_0 affect our outcome. In this instance we will definitely vary I_0 in increments larger than 0.0002.

MODEL REFINEMENT

If time allows, assessment and sensitivity analysis can lead to improvements in the model. Modeling, as pointed out earlier, is an iterative process, and refinements can almost always be made to develop more realistic scenarios. If the modeling is being done in a timed setting, such as for a competition or a homework assignment, then this may not be possible (although it is generally possible to indicate the type of refinements that would improve the model). However, for long-term projects, the model assessment really is an intermediate step before (possibly) starting the modeling loop over again. Discussing possible modifications to the model, even if you cannot make them, demonstrates that you are able to think beyond just the first approach. We demonstrate how this could be done with the variable-rate disease model.

In our previous disease model, the population was split into two classes: infected and not infected. If, however, we consider something such as an influenza outbreak, we know that people are capable of transmitting the flu for a short period of time, but eventually they develop immunity and will no longer spread the disease. This dynamic is certainly not captured with our initial disease models. With our new considerations in mind, we want infected individuals to be able to move out of the infected population after their bout with the

flu is over. It doesn't make sense to put them back into the susceptible population because we know that they have developed immunity. So we need to create another class, R , which represents those who have recovered from the disease.

We begin by defining the following:

P = the total number of people in the population,

S = the number of people in the population who are susceptible,

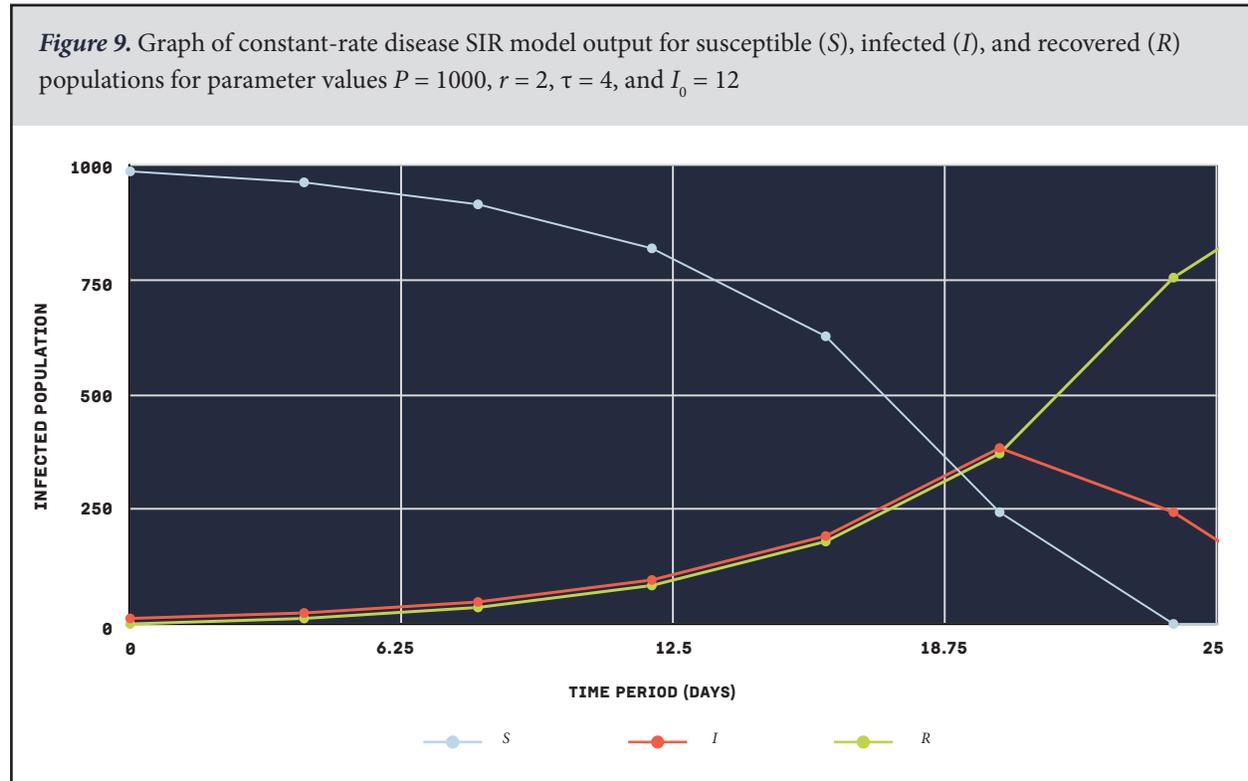
I = the number of people in the population who are infected and transmitting the disease,

R = the number of people in the population who have "recovered" (i.e., they are not susceptible and no longer transmit the disease).

Notice that for any time, $P = S + I + R$.

We will assume that each individual who becomes infected takes the same amount of time to recover. That is, our model will not take into account the possibility that one person recovers in 3 days while another takes 5 days to stop being a carrier of the disease (although a later refinement of our model might include a random distribution of recovery intervals).

6: ANALYSIS AND MODEL ASSESSMENT



Let's move forward looking at a specific example. Suppose at time $t = 0$, 12 people of a total population of 1000 become infected with a novel disease. We assume that each infected person recovers after exactly 4 days and that each infected person transmits the disease to 2 others during that 4-day period. So we will let $\tau = 4$ days.

At time $t = 0$, then, we have $I_0 = 12$, $S_0 = 988$, and $R_0 = 0$. After a 4-day time period, i.e., when $t = \tau$, all 12 of the infected individuals move to recovered status. In the meantime, they have infected 24 susceptible individuals. So we have

$$\begin{aligned} I(\tau) &= 2 \cdot I_0 = 2 \cdot 12 = 24, \\ S(\tau) &= 988 - I(\tau) = 988 - 24 = 64, \text{ and} \\ R(\tau) &= R_0 + I_0 = 0 + 12. \end{aligned}$$

We can continue as shown in Table 2 and Figure 9. Our refined model now shows a situation more consistent with our intuition regarding a disease such

as influenza; it starts out slowly, spikes, and then dies out as the remaining infected population recovers. This model is not perfect, but it does provide us with insight into the dynamic interaction of the affected populations.

Table 2. Infected, susceptible, and recovered populations over time for parameter values $P = 1000$, $r = 2$, $\tau = 4$, and $I_0 = 12$

t	$S(t)$	$I(t)$	$R(t)$	$P(t)$
0	988	12	0	1000
τ	964	24	12	1000
2τ	916	48	36	1000
3τ	820	96	84	1000
4τ	628	192	180	1000
5τ	244	384	372	1000
6τ	0	244	756	1000

IN SUMMARY

- 1 BE SURE TO ALLOCATE TIME TO ANALYZE YOUR RESULTS SINCE IT IS INDEED A CRITICAL PART OF THE ENTIRE MODELING PROCESS.
- 2 ALWAYS EXAMINE THE OUTPUT YOU GET FROM YOUR MODEL AND ASK YOURSELF IF IT MAKES SENSE. IF YOUR ANSWER DOESN'T MAKE SENSE, VERIFY THAT YOU HAVEN'T MADE A MISTAKE IN IMPLEMENTING YOUR MODEL.
- 3 IF YOUR SOLUTION IS CONSISTENT WITH YOUR ASSUMPTIONS BUT NOT CONSISTENT WITH THE REAL-WORLD PHENOMENON YOU ARE TRYING TO DESCRIBE, YOU MAY NEED TO REFINE YOUR MODEL BY ADJUSTING YOUR ASSUMPTIONS.
- 4 LIST STRENGTHS AND WEAKNESSES OF YOUR MODEL.
- 5 TRY TO DETERMINE HOW SENSITIVE YOUR MODEL IS TO PARAMETERS AND ASSUMPTIONS.
- 6 INCLUDE SPECIFIC IMPROVEMENTS YOU WOULD HAVE INCORPORATED GIVEN MORE TIME.

ACTIVITY

READ THE SOLUTION TO THE RECYCLING PROBLEM. WHAT ARE THE STRENGTHS AND LIMITATIONS OF THAT MODEL? WHAT ARE PARAMETERS IN THE RECYCLING MODEL THAT COULD BE EXAMINED FOR SENSITIVITY? HOW MIGHT THE SENSITIVITY OF THE RECYCLING MODEL BE ANALYZED? WRITE A FEW PARAGRAPHS ASSESSING THIS MODEL.

7. PUTTING IT ALL TOGETHER

Now that a model has been created, solved, and assessed, the time comes to write everything up as a polished solution paper. This step is just as important as the effort necessary to get to this point. Keep in mind that you are the expert about the problem, and now your role is to explain what you did in detail to people unfamiliar with your solution approach. To this end, it is critical that you take good notes from the initial brainstorming process through the final analysis to be sure you have kept track of all the assumptions you made. Good writing also takes time, so be sure to allocate a period of time to step back from the math modeling and focus on quality writing. In this section we discuss how to structure your report and some key points for successful technical writing.

This step is just as important as the effort necessary to get to this point. Keep in mind that you are the expert about the problem.

STRUCTURE

A technical report typically starts with a **summary page**, also called an *executive summary* or an *abstract*, that is of one page or shorter. This is not an introduction; it's actually a place to summarize how the problem was solved and to provide a brief description of the results. It might seem strange to put the conclusion at the beginning, but this “bottom-line up front” approach is convenient for those reading your report.

The abstract or summary page should restate the problem, briefly describe the chosen solution methods, and provide the final results and conclusions. You should describe your results in complete sentences that can stand on its own, without using variables. The summary lets the reader know what to expect in the report but does not overwhelm him or her with unfamiliar mathematical notation. Imagine that a reader will decide whether to continue reading the rest of your paper itself based on this abstract. As an example, see the solution to the recycling problem.

After the summary, the paper should include a **formal introduction** that includes a restatement of the underlying real-world application as if the reader does not have any prior knowledge. This section usually contains some motivation or relevant background information as well but should not include a lengthy history lesson. Both the general modeling question as well as the **concise problem statement** that you developed should be at the forefront of this section. This section must provide a paragraph that describes how you approached the problem. For example, if we consider the task of ranking roller coasters based on how thrilling they are, then it would help to define “thrilling” up front. For example, “Our model is based on the notion that rides with high accelerations, inversions, and significant heights are thrilling.” Note that this statement doesn't exactly explain how those features are quantified or implemented for a mathematical ranking system for thrilling roller

coasters, but it does give the reader an idea of what will be showing up later in the document. It may seem counterintuitive, but the introduction and abstract are typically written last. This is because, after all else has been written, the author has a complete picture of the manuscript and may then best tailor these sections accordingly.

The **body of the solution paper** will likely be several pages long and split up into sections about assumptions, variables, the model, the solution process, analysis, and overall conclusions. Let the reader know about the overarching assumptions you made to make the problem solvable. Some specific assumptions may need to be included again later, within the paper's main text, in order to clarify certain ideas as they are developed. However, the most important message here is that all assumptions are included and listed at some point in your write-up. You should be sure to justify why those assumptions are reasonable and include citations as needed. Plagiarism of any kind is never acceptable.

When you next begin to describe your model and how you solved it, clearly state the variables you will be using and the corresponding units. If there are relationships between variables, explain where they come from and, if needed, refer to any necessary assumptions. Mathematical equations and formulas should be centered, each occurring on their own line. We provide some more specific details on this in the following section.

Finally, the paper must have a **conclusion** section that recaps the important features of the model. It is critical that this section includes an analysis of your results, as described in the previous section. An honest assessment of the strengths and weaknesses is important. In particular, you can comment on how the model can be verified and how sensitive the model is to the assumptions. We proceed by giving some tips about technical writing.

TECHNICAL WRITING DOS & DON'TS

1. DO NOT WRITE A BOOK REPORT.

It is critical that the narrative of your solution doesn't read like a story about how you came up with your model. For example, consider the following two excerpts about an assumption made for the recycling model.

Example 1:

A study of drop-off recycling participation in Ohio found that 15.5% of citizens who do not have access to curbside recycling use drop-off recycling [8]. We have assumed that this data is representative of the U.S.

Example 2:

We were stuck because we did not know how many people in the U.S. recycle. We googled and found an article that Ohio's participation in drop-off recycling was 15.5% for people who did not have access to curbside recycling, so we used that number in our model.

The second example is written in a way that makes the assumption sound invalid or that it was chosen only because no other information could be found. However, the first one sounds as though some research was done and a useful and legitimate source was identified, which provided an applicable statistic.

2. DO NOT USE WORDS WHEN USING MATHEMATICS WOULD BE MORE APPROPRIATE.

Which of the following is more effective?

Example 1:

For an ideal gas, we have

$$P = \frac{nRT}{V},$$

where P is the absolute pressure of the gas, V is the volume of the gas, n is the amount of substance of gas (measured in moles), T is the absolute temperature of the gas, and R is the ideal, or universal, gas constant.

Example 2:

For an ideal gas, the absolute pressure is directly proportional to the product of the number of moles of the gas and the absolute temperature of the gas and inversely proportional to the volume of the gas, with proportionality constant R , the ideal, or universal, gas constant.

In this case, the mathematics is easier to follow, and you can imagine that the more complicated the calculations, the harder it would be to try to describe in words.

3. DO USE PROPER SENTENCE STRUCTURE WHEN EXPLAINING MATHEMATICS.

Communicating mathematics requires proper punctuation, such as periods at the end of a computation if the computation ends the sentence, as in Example 1 below. Use commas when appropriate.

Example 1:

If s is the length of the side of the square box, then the area of a side is given by

$$A = s^2,$$

and the volume is given by

$$V = s^3.$$

Example 2:

If s is the length of the side of a square box then we can find the area and volume.

$$A = s^2$$

$$V = s^3$$

4. DO NOT SUBSTITUTE MATHEMATICAL SYMBOLS FOR WORDS WITHIN SENTENCES, AS IN THE SECOND OF THE FOLLOWING TWO EXAMPLES.

Example 1:

For this work L is the length of the side of a rectangle.

Example 2:

For this work $L =$ the length of the side of the rectangle.

5. DO PAY ATTENTION TO SIGNIFICANT FIGURES.

For example, your calculator might read a value of 27.3416927482, but you may not need to report all of those digits unless you are trying to show accuracy in the later decimal places.

6. DO USE SCIENTIFIC NOTATION WHEN NUMBERS VARY BY ORDERS OF MAGNITUDE,

meaning that the exponent is really what matters in understanding the significance of the value. For example, the diameter of the sun is $1.391e^6$ km, while the diameter of a baseball is $2.290e^{-4}$ km.

7: PUTTING IT ALL TOGETHER

7. DO LABEL FIGURES

and use a large enough font so that the axes are clearly readable.

8. DO NOT FORGET TO INCLUDE UNITS AS APPROPRIATE.

9. DO CHECK CAREFULLY FOR SPELLING AND GRAMMAR MISTAKES,

especially those that spell check might miss. For example, it's easy to confuse *their*, *there*, and *they're*.

10. DO GIVE CREDIT WHERE CREDIT IS DUE.

This means including citations and building your bibliography as you go.

Technical writing takes practice, but the end result should be something of which to be extremely proud. In reviewing your final paper, you can step back and look at all you have accomplished throughout the modeling process. Ultimately your model can lead to the creation of new knowledge and provide deeper insight about the world we live in. Remember that modeling also takes practice so the next time you tackle an open-ended problem you will already have a new set of tools that will make the entire experience go more smoothly.

IN SUMMARY

- 1 TAKE NOTES THROUGHOUT YOUR ENTIRE MODELING PROCESS SO THAT YOU DO NOT LEAVE OUT ANYTHING IMPORTANT, ESPECIALLY ASSUMPTIONS MADE ALONG THE WAY.
- 2 GIVE YOURSELF ENOUGH TIME TO FOCUS ON THE WRITING PROCESS AND TO PROOFREAD THE REPORT.
- 3 KEEP IN MIND THAT THIS IS A TECHNICAL DOCUMENT, NOT A STORY ABOUT YOUR MODELING EXPERIENCE.
- 4 FOLLOW THE GUIDELINES FOR TECHNICAL WRITING.
- 5 SOME ADDITIONAL REFERENCES ON TECHNICAL WRITING CAN BE FOUND AT [3].
- 6 PAT YOURSELF ON THE BACK FOR YOUR ACCOMPLISHMENTS.

APPENDICES & REFERENCES

A. FORWARD EULER METHOD

Let's explore how the forward Euler method works in the context of the varying-rate disease model.

As a reminder, our model is described with the differential equation

$$\frac{dI}{dt} = kI(t)(P - I(t)),$$

where I is the number of infected individuals, P is the total population, and k is a positive constant. We'll use the same parameter values we used to plot the analytic solution: $k = 0.0006$, $P = 1000$, and $I_0 = 20$.

The method we will use here is called the forward Euler method, which takes advantage of the fact that the derivative is the same thing as slope of the tangent line. We'll demonstrate the method on this example, but do not want to overwhelm the reader with too many details. We point you toward [9, 4] for a deeper look into this very powerful approximation method.

We start with the initial population, $I_0 = 20$. That is, when $t = 0$ days, $I = 20$. In other words, we know that the point $(0, 20)$ is on the graph of the solution. We also know what the slope of the solution curve is at that point because we have an equation for $\frac{dI}{dt}$:

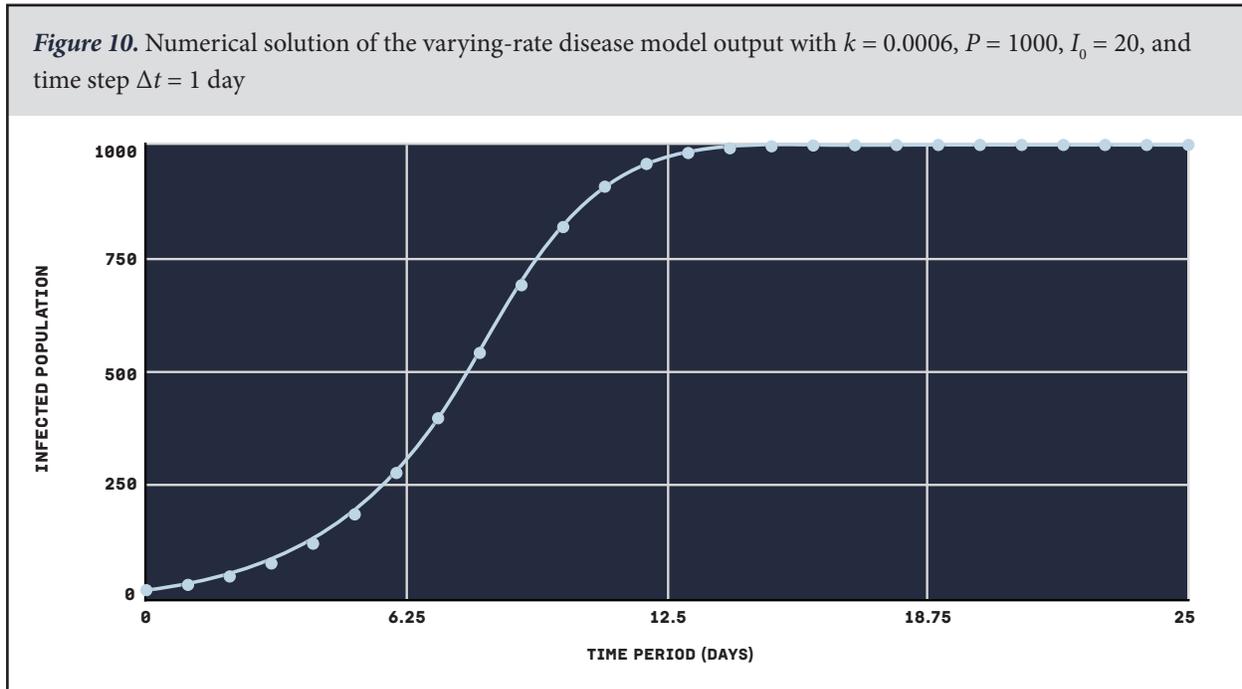
$$\begin{aligned} \left. \frac{dI}{dt} \right|_{t=0} &= kI(0)(P - I(0)) \\ &= kI_0(P - I_0) \\ &= (0.0006) \cdot 20 \cdot (1000 - 20) \\ &= 11.76. \end{aligned}$$

Now we can write the equation of the line through the point $(0, 20)$ with slope of 11.76 as follows. Recall that the independent variable is t and the dependent variable is I .

$$\begin{aligned} I - 20 &= 11.76(t - 0), \\ I &= 11.76(t - 0) + 20. \end{aligned}$$

APPENDICES & REFERENCES

Figure 10. Numerical solution of the varying-rate disease model output with $k = 0.0006$, $P = 1000$, $I_0 = 20$, and time step $\Delta t = 1$ day



Hence, we have a line with slope of 11.76 and y -intercept (or I -intercept, in this case) of 20.

We can use this linear function as an approximation for the true solution. We assume that the solution to the differential equation is approximately the same as this line for points nearby. Perhaps we'll assume it's a good enough approximation through time $t = 1$. We can find the number of infected individuals at $t = 1$.

$$I(1) = 11.76(1 - 0) + 20 = 31.76.$$

(Note: While it's not possible to have a fractional person—or, more specifically, 0.76 of an infected individual—it doesn't deter us from continuing to approximate the solution using this method. We do suggest noting that something unrealistic has occurred and encourage you to re-examine it later when assessing your model.) Thus we have a line segment from $(0, 20)$ to $(1, 31.76)$.

Now you can image us starting the process over. In other words, we assume that the point $(1, 31.76)$ is on the solution curve, and we can use the derivative to give us the slope at that point:

$$\begin{aligned} \left. \frac{dI}{dt} \right|_{t=1} &= k I(1) (P - I(1)) \\ &= (0.0006) \cdot 31.76 \cdot (1000 - 31.76) \\ &= 18.45. \end{aligned}$$

As before, we can find the equation of the line through the point $(1, 31.76)$ with slope of 18.45.

$$\begin{aligned} I - 31.76 &= 18.45(t - 1) \\ I &= 18.45(t - 1) + 31.76. \end{aligned}$$

We will assume that this makes a good enough approximation for the solution through $t = 2$. Thus we estimate the population at time $t = 2$ to be

$$I(2) = 18.45(2 - 1) + 31.76 = 50.21.$$

Once again, we have identified a solution method requiring multiple iterative calculations, which can be easily performed using technology such as Excel.

As before, now that we have a table of values in Excel, we can generate a plot of our numerical solution. (See Figure 10.)

B. THE 2013 M³ CHALLENGE PROBLEM AND THE SOLUTION PAPER FROM TEAM 1356

Waste Not, Want Not: Putting Recyclables in Their Place

Plastics are embedded in a myriad of modern-day products, from pens, cell phones, and storage containers to car parts, artificial limbs, and medical instruments; unfortunately, there are long-term costs associated with these advances. Plastics do not biodegrade easily. There is a region of the Northern Pacific Ocean, estimated to be roughly the size of Texas, where plastics collect to form an island and cause serious environmental impact. While this is an international problem, in the U.S. we also worry about plastics that end up in landfills and may stay there for hundreds of years. To gain some perspective on the severity of the problem, the first plastic bottle was introduced in 1975 and now, according to some sources, roughly 50 million plastic water bottles end up in U.S. landfills every day.

The United States Environmental Protection Agency (EPA) has asked your team to use mathematical modeling to investigate this problem.

How big is the problem? Create a model for the amount of plastic that ends up in landfills in the United States. Predict the production rate of plastic waste over time and predict the amount of plastic waste present in landfills 10 years from today.

Making the right choice on a local scale. Plastics aren't the only problem. So many of the materials we dispose of can be recycled. Develop a mathematical model that a city can use to determine which recycling methods it should adopt. You may consider, but are not limited to:

- providing locations where one can drop off pre-sorted recyclables
- providing single-stream curbside recycling
- providing single-stream curbside recycling in addition to having residents pay for each container of garbage collected

Your model should be developed independent of current recycling practices in the city and should include some information about the city of interest and some information about the recycling method. Demonstrate how your model works by applying it to each of the following cities: Fargo, North Dakota; Price, Utah; Wichita, Kansas.

How does this extend to the national scale? Now that you have applied your model to cities of varying sizes and geographic locations, consider ways that your model can inform the EPA about the feasibility of recycling guidelines and/or standards to govern all states and townships in the U.S. What recommendations does your model support? Cite any data used to support your conclusions.

Submit your findings in the form of a report for the EPA.

The following references may help you get started:

<http://www.epa.gov/epawaste/nonhaz/municipal/index.htm>

http://5gyres.org/what_is_the_issue/the_problem/

Moody's Mega Math Challenge supports
Mathematics of Planet Earth (MPE2013).
www.mpe2013.org



Analysis of Plastic Waste Production and Recycling Methods

EXECUTIVE SUMMARY

In 2010 alone, the U.S. generated approximately 250 million tons of trash [1]. Much of this waste consisted of plastics, which build up in landfills and flow into oceans through storm drains and watersheds [2], breaking up into little pieces and absorbing contaminants in the process. A major method to reduce waste is recycling, where materials like glass, paper, and plastic are reformed to create new products. There are many different methods of collection of recyclable materials, including drop-off centers, where citizens transport their recyclables; single stream curbside collection, where the city collects the recyclables of each household; and dual stream curbside collection, where the city collects recyclables that are pre-sorted by each household. To encourage or subsidize recycling programs, some cities may implement a Pay-As-You-Throw (PAYT) program, where citizens pay a fee based on the amount of garbage they throw away.

The EPA tasked us to analyze the production and discard rate of plastic waste over time. We were also asked to create a model of possible methods for recycling collection to determine which methods are appropriate for what cities. Using a linear regression model over years passed since 2000, we estimated that 35.1 million tons of plastic waste will be discarded in 2023. We also modeled the use of drop-off centers, single stream curbside collection, and dual stream curbside collection to calculate the total amount of recyclables collected as well as the cost to the city using each recycling method.

For collection using drop-off centers, we developed a simulation that randomly simulated the number of households who would recycle when drop-off centers were placed around the city. The simulation took into account the area, population, average household, maximum distance citizens are willing to travel, and number of drop-off centers. Using these data, we calculated the amount recycled and then calculated the net cost to the city by subtracting operating costs of Material Recovery Facilities (MRFs) from revenue generated by selling recycled products.

For curbside collection, we calculated the number of trucks needed to service a given city, based on population density. Based on labor, upkeep, and fuel, we calculated the costs of a curbside collection program. Again, using the calculated amount of collected material, we determined the net revenue generated by these products.

We determined that by using a drop-off center method, Fargo and Wichita would generate profits, while Price would incur costs that could be partially covered by using Pay-As-You-Throw. Using a curbside collection method, Fargo and Price would incur costs that could be partially covered by Pay-As-You-Throw, while Wichita would generate profits using either single or dual stream collection. Thus, either drop-off or curbside collection methods may be feasibly implemented in cities around the U.S., depending on population and area of each city. We concluded that small cities tend to incur net costs from recycling programs, while larger cities like Wichita may profit from using a dual stream curbside collection program.

To assess use of recycling programs on a national level, we programmed a computer simulation generating an image of all the counties of the U.S., where blue dots on the U.S. map represented counties where at least one of our three proposed recycling programs earned a net profit. In general, we recommend that the EPA extend recycling program guidelines to the national level.

I. INTRODUCTION

1. Background

Each year, the U.S. consumes billions of bags and bottles. However, of the plastics that the U.S. produces, only 5% is recovered [2]. Unrecycled plastics present a growing hazard because they contain dangerous chemicals like polycarbonate, polystyrene, PETE, LDPE, HDPE, and polypropylene, which accumulate over time and build up in our oceans and landfills. As such, it is important to assess the scale of our plastic waste production problem over time.

Our foremost method of reducing wastes like plastics is through recycling, where useful materials including glass, plastic, paper, and metals are recovered so that they may be used to create new products [3]. There exist several methods of recycling collection; in general, cities may use either use drop-off centers or curbside collection. With drop-off centers, the residents carry the burden of transporting their recyclable waste, while curbside collection places this burden on the city. If a city implements curbside collection, it may choose to use single stream, dual stream, or pre-sorted methods; in single stream, all recyclables are collected as one unit, whereas in dual stream, recyclables are separated into paper and glass, cans, and plastic [4]. Further separation exists with the pre-sorted collection method, where recyclables are fully separated by material type [5]. There are advantages and disadvantages associated with each method of collection, and in choosing the type of recycling program to implement, cities must consider, among other factors, the practicality of individual household collection, as well as the volume of recyclables that would be collected using each program [6]. Some communities may use Pay-As-You-Throw (PAYT) programs, which encourage residents to recycle their waste so as to avoid fees dependent on the weight of their trash [7]. We assess in this analysis whether it is more efficient to use drop-off centers or curbside collection, depending on the city where the recycling program is being implemented, as well as the effect of using PAYT programs to generate additional revenue for the city.

2. Restatement of the Problem

In this analysis, we were requested by the EPA to create a model to predict the change in plastic production rate over time, as well as the amount of plastic waste in landfills in the year 2023. We were further asked to look at various recycling methods, not limited to the recycling of plastics, and to analyze the recycling method a city should develop, using as sample points the cities of Fargo, North Dakota; Price, Utah; and Wichita, Kansas. Finally, the EPA requested that we provide recommendations for developing recycling methods on the national level based on the model we designed.

3. Global Assumptions

Throughout our analysis, we will make the following assumptions:

- 1 A city's population is approximately evenly distributed. Population mostly varies on a large scale: in the small microcosm of a city, the population density will not vary much.
- 2 A city's shape is approximately square. Most cities are shaped like this, as are the three sample cities we were provided with.
- 3 A city's roads are laid out in a grid plan. The popularity of the grid plan is pervasive, dating back to Ancient Rome, and most cities are organized as such, like our three sample cities. [8, 9, 10].

- 4 A household’s recycling stance is consistent. That is, a household that recycles always recycles, and a household that does not will never recycle. Recycling is a habit, and households that recycle tend to recycle consistently.

II. ANALYSIS OF THE PROBLEM AND THE MODEL

1. Plastic Waste Production

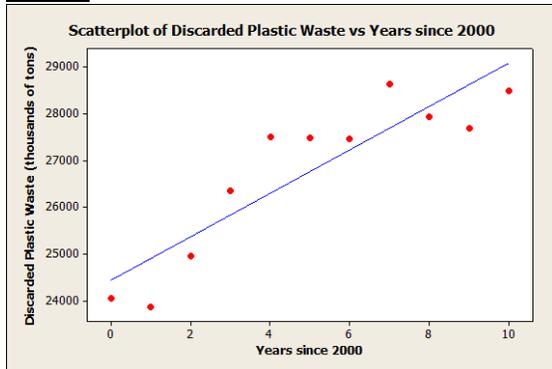
Assumptions

- 1 We used data collected from the past ten years because the first plastic bottle was introduced in 1975 [11], and recycling has only become important recently. In other words, values used before 2000 would not adequately take into account the recycling methods which have now become widespread.

Approach

We created our model by performing linear and logistic regressions on the amount of plastic waste discarded per year for the last decade in thousands of tons as provided by the EPA [1].

Model

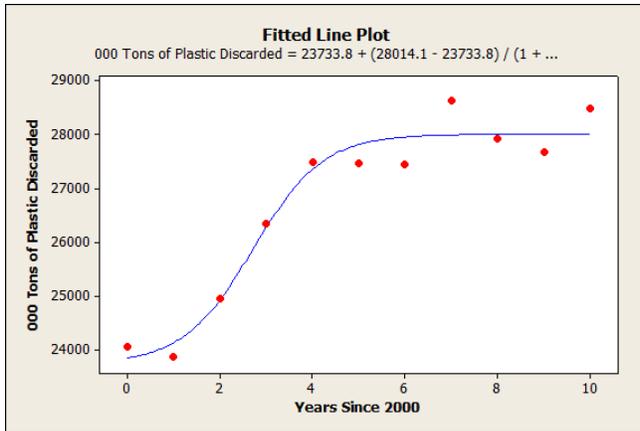


$$\text{Discarded Plastic Waste (thousands of tons)} = 463.27 * (\text{years since 2000}) + 24443.6$$

$$R^2 = .803; S = 801$$

The R^2 value of .803 means that 80.3% of the variability in the amount of plastic waste discarded is explained by the linear relationship between years passed since 2000 and plastic waste amount. The standard deviation of the residuals was 801.

Based on this model, the amount of unrecycled plastic waste discarded in 2023 will be $463.27 * 23 + 24443.6 = 35098.81$ thousands of tons, or 35.1 million tons.



$$\text{Discarded Plastic Waste (thousands of tons)} = 23733.8 + (28014.1 - 23733.8) / (1 + \exp((\text{Years since 2000} - 2.71325) / -0.752611))$$

$$S = 422$$

The previously mentioned R^2 value only makes sense under the assumption that the linear model was appropriate. Since there is a prominent bend in the data, we fit them with a logistic curve as well. A statistical software found the four parameters using a successive approximation method, and produced the model above.

The standard deviation of the residuals in this model is only 422, which is almost twice as small as it was in the linear model. Unfortunately, this model assumes that the tonnage of discarded plastic waste will level off, which is not entirely reasonable. It does, however, give a best-case result (e.g. if recycling initiatives work perfectly). The projected value of discarded waste for 2023 is 28014.1 thousand tons (within 4 decimal places), which is the maximal value according to the model.

In summary, the linear model (which seems to overpredict the later values) yields a value of 35.1 million tons, while the logistic model (which levels off) predicts that it will level off at 28.0 million tons. The US population has been increasing linearly since 2000 [12], so the linear model gives a more plausible value for the next ten years.

2. Recycling Methods

Assumptions

- 1 City shape can be approximated as a square or diamond. Most cities in the U.S. are square-like in shape, including Fargo, North Dakota, Price, Utah, and Wichita, Kansas.
- 2 The streets of the city are laid out in a grid. Many large cities have streets following a grid, including Fargo, North Dakota; Price, Utah; and Wichita, Kansas all use grid systems
- 3 There is no overlap between use of drop-off centers and curbside collection.
- 4 The composition of recyclables in the MSW stream is fixed over the entire planning horizon.

Model 1: Drop-off Centers

Approach

Assumptions

- 1 Each household makes a collective decision on whether or not to recycle because it is convenient for a household to transport all of their recyclables to a drop-off center together.
- 2 The probability of a household's deciding to recycle varies linearly with the household's distance to the nearest drop-off center.
- 3 Recycling households recycle all recyclable waste.

To assess the amount of recyclables collected by a recycling program dependent on drop-off centers, we created a computer simulation where we assumed uniform population density and where we place equally spread drop-off centers around the city, as many as would fit without overlapping coverage. To determine whether each household would recycle, a random number from 0 to 1 is generated, and if the number is less than the household's probability of recycling, which we assumed varies linearly with the household's distance to the nearest drop-off center, the household recycles. We also determined the cost of maintaining each recycling center and the revenue the center would generate, and used these data to calculate the total cost to the city of the drop-off center program. In our simulation, we accepted as inputs the area of the city, population of the city, average number of people in a household, maximum distance citizens are willing to travel, and number of drop-off centers.

Taxicab Distance

Because streets are assumed to be organized in a grid, we calculate distance as "taxicab distance", or distance in which the only path allowed consists of horizontal and vertical lines. In other words, given p_x and p_y as the coordinates of the drop-off center, and x and y as the coordinates of the household, the distance between them, d , can be calculated as:

$$d = |y - p_y| + |x - p_x|$$

A Household's Maximum Distance to a Drop-off Center

Assumption

- 1 Recycling households make biweekly trips to a drop-off center.

We recommend that cities conduct a survey to determine the distance their citizens are willing to travel in order to recycle, though we calculated this distance in our model. U.S. citizens are willing for their household to pay \$2.29 a month for curbside collection [13]. Since this is the amount that they are willing to pay to recycle at greatest convenience, we can assume that it is equivalent to the maximum amount they are willing to pay as the driving cost to a drop-off center.

The average price of a gallon of gas is \$3.784 [14] and the average mileage of a passenger car in 2010 was 23.8 mpg [15]. The cost of traveling a distance d is:

$$\text{Cost} = (\$3.784/\text{gallon}) * d / (23.8 \text{ miles/gallon})$$

The distance citizens are willing to travel each week is:

$$(\$2.29 \text{ dollars/month}) / (4.35 \text{ weeks/month}) = \$0.53 \text{ dollars/week} = (\$3.784/\text{gallon}) * (d \text{ miles/week}) / (23.8 \text{ miles/gallon})$$

$$d \text{ miles/week} = (\$0.53 \text{ dollars/week}) * (23.8 \text{ miles/gallon}) / (\$3.784 \text{ dollars/gallon}) = 3.33 \text{ miles/week}$$

Since citizens must drive to the drop-off center and back, the maximum distance driven to the drop-off center is 1.665 miles/week. Assuming that households make biweekly recycling trips, the maximum distance from a household to a drop-off center for the household to consider recycling is 1.665 miles/week * 2 weeks = 3.33 miles.

A study of drop-off recycling participation in Ohio supports our model, finding that the functional usage area of a full-time urban drop-off center is about 3.5 miles [16].

Number of Recycling Households Covered by a Drop-off Center
Assumption

- 1 The available data from Ohio are representative of the U.S. as a whole.

Each drop-off center will receive recyclables from households up to 3.33 miles away. Using taxicab distance, which only allows horizontal and vertical movement, the area within 3.33 miles is bounded by a diamond (a square rotated 45°). The diagonal of the diamond is twice the distance from the center to a corner, or 2 * 3.33 miles = 6.66 miles. Since the diamond is a square, diagonal length = square root(2) * side length, so the side length is 4.71 miles. The area of the diamond is side length ^ 2 = 22.18 sq. mi. This is the coverage area of the drop-off center, which contains all the households that will consider using the drop-off center.

The number of households in the drop-off center’s coverage area is:

$$\text{Households} = 22.18 \text{ sq. mi} * (\text{population} / \text{land area}) / (\text{average household size})$$

A study of drop-off recycling participation in Ohio found that 15.5% of citizens who do not have access to curbside recycling use drop-off recycling [16]. Assuming that this data is representative of the U.S. as a whole, the number of recycling households covered by each drop-off center is:

$$\text{Recycling households} = 22.18 \text{ sq. mi} * (\text{population} / \text{land area}) / (\text{average household size}) * .155$$

In our simulation, we assigned 15.5% as the median household probability of recycling. The closer a household is to the drop-off center, the more likely it is to recycle. Within the drop-off center coverage area, the closer half of households has a greater than 15.5% recycling probability and the farther away half of households has a less than 15.5% recycling probability. The distance from the center to the boundary of the closer half of households is the diagonal of the square with half the area of the entire coverage area, which is:

$$\text{Halfway distance} = \text{square root}(22.18 \text{ sq. mi.} / 2) * \text{square root}(2) = 4.71 \text{ miles}$$

We assumed that the probability of a household recycling varies linearly with the distance to the nearest recycling center. At a distance of 4.71 miles, the probability is 15.5%. At the boundary distance of 6.66 miles, the probability is 0%. Extending the line through these points, at the center, with a distance of 0 miles, the probability is 52.9%. In our simulation, the number of recycling households covered by each drop-off center is approximately the same as that calculated using the formula previously given.

Drop-off Center Placement

In our simulation, we placed as many drop-off centers as possible in each city so that none of the coverage areas overlap, with at least one drop-off center in each city. The cost efficiency of drop-off centers decreases when their coverage areas overlap.

Annual Amount Recycled

The average American generates 4.5 pounds of waste per day [17], about 75% of which is recyclable [18]. Thus, the average American generates 4.5 pounds * 0.75 = 3.375 pounds of recyclable waste per day.

$$\text{Annual Amount Recycled (tons)} = (\text{recycling households}) * (\text{average household size}) * 3.375 \text{ lb} * 365 \text{ days/year} * 0.005 \text{ lb/ton} * (\# \text{ drop-off centers})$$

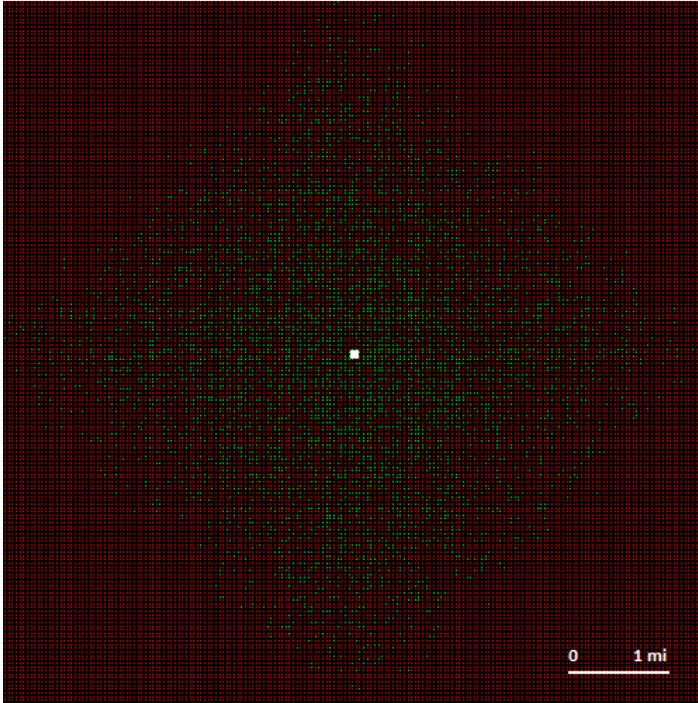
This formula can be used in place of our simulation to calculate annual amount recycled, as long as there is no overlap between drop-off center coverage areas and the drop-off center coverage area is entirely contained within the city. For example, because the drop-off center coverage area (22.18 sq. mi.) is much larger than the area of Price, Utah (4.2 sq. mi.), this formula cannot be used in place of our simulation for Price, Utah.

Using our simulation, we were able to calculate the annual amount recycled for Fargo, North Dakota; Price, Utah; and Wichita, Kansas, as well as to visualize the households contributing recyclables to each city. In the screenshots below, the white dots represent the drop-off centers; the green dots represent households that are recycling; and the red dots represent households that are not recycling.

Fargo, North Dakota

Annual Amount Recycled (tons) = 5209.66

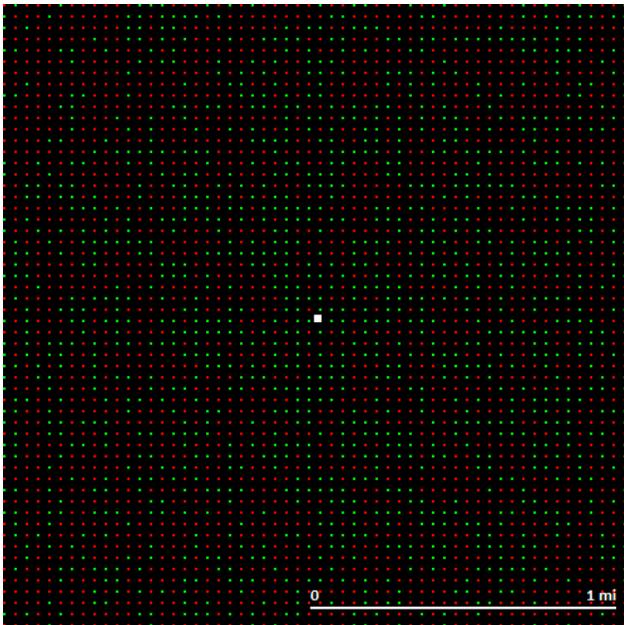
Number of people recycling= 8458



Land Area = 48.82 sq. mi.
Population = 105,549 people
Average household size = 2.15 people

Price, Utah

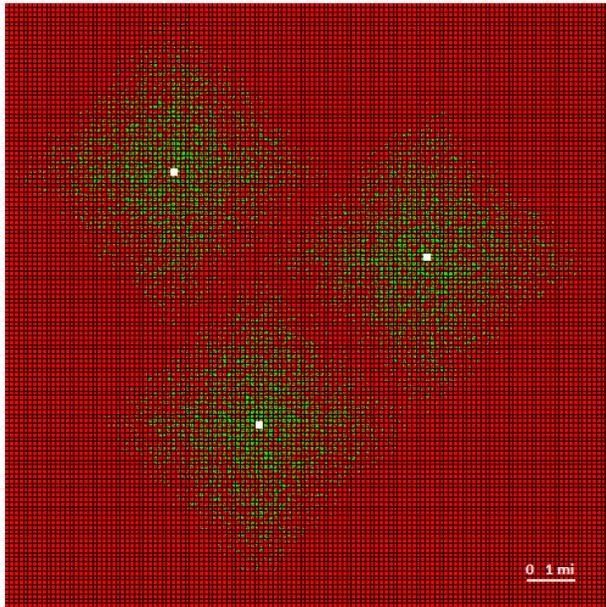
Annual Amount Recycled (tons) = 1876.88
Number of people who recycle = 3047



Land Area = 4.2 sq. mi.
 Population = 8,402 people
 Average household size = 2.60 people

Wichita, Kansas

Annual Amount Recycled (tons) = 16929.56
 Number of people who recycle = 27486



Land Area = 159.29 sq. mi.
 Population = 382,368 people
 Average household size = 2.48 people

Drop-off Center Cost

A report by design engineering company R.W. Beck, Inc. recommends front load dumpsters as the most cost-effective type of drop-off center. Under this plan, front load dumpsters would be set up at each drop-off center site and recyclables would be collected in two streams, commingled containers and paper. The annual cost of a front load dumpster site is about \$5,575 per year [19]. Thus, the total annual cost of drop-off centers is:

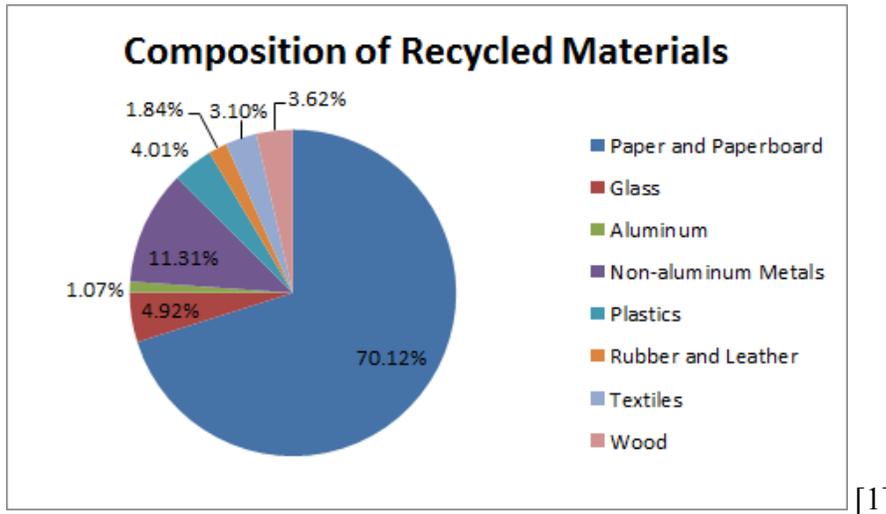
$$\text{Annual cost of drop-off centers} = \$5,575 * (\# \text{ drop-off centers})$$

Revenue Generated

To calculate the total revenue per ton generated from selling recycled products, we used the following formula, taking into account the market price per ton for each product [20, 21, 22, 23, 24, 25, 26, 27]:

$$\text{Revenue per ton} = \text{Revenue}_{\text{Paper}} + \text{Revenue}_{\text{Glass}} + \text{Revenue}_{\text{Ferrous Metals}} + \text{Revenue}_{\text{Aluminum}} + \text{Revenue}_{\text{Plastic}} + \text{Revenue}_{\text{Textiles}} + \text{Revenue}_{\text{Wood}} = (.7012 * \$112.82) + (.0492 * \$13) +$$

$$(.1131 * \$217.75) + (.0107 * \$310) + (.0401 * \$370) + (.031 * \$100) + (.0362 * \$296) + (\$135 * .0180) = \$128.78 \text{ per ton of recycled material}$$



Based on a study conducted on recycling collection and processing options in New Hampshire [28], cities can decide between small, medium, and large Materials Recovery Facilities (MRFs) depending on the annual tonnage. The cost per ton using dual stream and cost per ton using single stream varies depending on the size of the MRF. For drop-off centers, we are assuming that dual stream is used.

Fargo, North Dakota

We calculated that Fargo would collect 5,209.66 tons of recyclables. This suggests that a medium tonnage mini MRF, which has an annual tonnage of 5,283, is sufficient for the city. The cost per ton of a medium mini MRF using dual stream is \$124.62. Since the material revenue per ton was previously found to be \$128.78, we can calculate the net cost per ton as:

$$\text{Net cost} = \$124.62 - \$128.78 = -\$4.16$$

The total cost to the city can then be calculated as:

$$\text{Total cost} = -\$4.16 \text{ per ton} * 5,209.66 \text{ tons} + \$5,575 \text{ per drop-off container} * 1 \text{ container} = -\$16,097.19 \text{ (profit)}$$

Price, Utah

We calculated that Price would collect 1876.88 tons of recyclables. Price would use a low tonnage mini MRF, and the net cost per ton would also be \$89.69. Then, the total cost to the city is:

$$\text{Total cost} = \$89.69 \text{ per ton} * 1876.88 \text{ tons} + \$5,575 \text{ per drop-off container} * 1 \text{ container} = \$173,912.37$$

Wichita, Kansas

We calculated that Wichita would collect 16,929.56 tons of recyclables, suggesting that Wichita would require a high tonnage mini MRF, which has an annual tonnage of around 7,500. For a high tonnage mini MRF, the cost per ton for dual stream is \$95.40. Since the material revenue is \$128.78 per ton, the net cost per ton is:

$$\text{Net cost} = \$95.40 - 128.78 = -33.38$$

This represents a profit of \$33.38 per ton of recycled material. The total cost to the city is then:

$$\text{Total cost} = -\$33.38 \text{ per ton} * 16,929.56 \text{ tons} + \$5,575 \text{ per drop-off container} * 1 \text{ container} = -\$559,533.71 \text{ (profit)}$$

Pay-As-You-Throw

If the city implements a Pay-As-You-Throw (PAYT) program, it will collect revenue from citizens who must pay an amount depending on the volume of waste they generate. We can calculate revenue generated by such a program by using the formula [29]:

$$\text{Revenue}_{\text{PAYT}} = (\text{Weight}_{\text{Waste}} / \text{Volume}_{\text{Container}} * \text{Price}_{\text{Container}} - \text{Price}_{\text{Startup, Maintenance per day}}) * \text{Population}$$

The average American generates 4.5 lbs of waste and recycles 1.5 lbs [17]. PAYT programs cost around \$0.28 per capita, based on surveys of Wisconsin and Iowa [30]. We also simplified $\text{Volume}_{\text{Container}} * \text{Price}_{\text{Container}}$ as Container price/pound, since the containers are meant to hold specific amounts of weight. Thus, the revenue generated by PAYT for citizens who recycle can be calculated as:

$$\text{Revenue}_{\text{PAYT, Recycle}} = ((4.5 \text{ lbs} - 3.375 \text{ lbs}) * \text{Container price/pound} - \$0.28/365) * \text{Population}_{\text{Recycle}}$$

$$\text{Revenue}_{\text{PAYT, Don't recycle}} = (4.5 \text{ lbs} * \text{Container price/pound} - \$0.28/365) * \text{Population}_{\text{Don't recycle}}$$

$$\text{Total revenue}_{\text{PAYT}} = \text{Revenue}_{\text{PAYT, Recycle}} + \text{Revenue}_{\text{PAYT, Don't recycle}} = ((4.5 \text{ lbs} - 3.375 \text{ lbs}) * \text{Container price/pound} - \$0.28/365) * \text{Population}_{\text{Recycle}} + (4.5 \text{ lbs} * \text{Container price/pound} - \$0.28/365) * \text{Population}_{\text{Don't recycle}}$$

Fargo, North Dakota

$$\text{Total revenue}_{\text{PAYT}} = (1.125 \text{ lbs} * x - \$0.28/365) * 8,458 \text{ people}_{\text{Recycle}} + (4.5 \text{ lbs} * x - \$0.28/365) * (105,549 - 8,458 \text{ people}_{\text{Don't recycle}})$$

Price, Utah

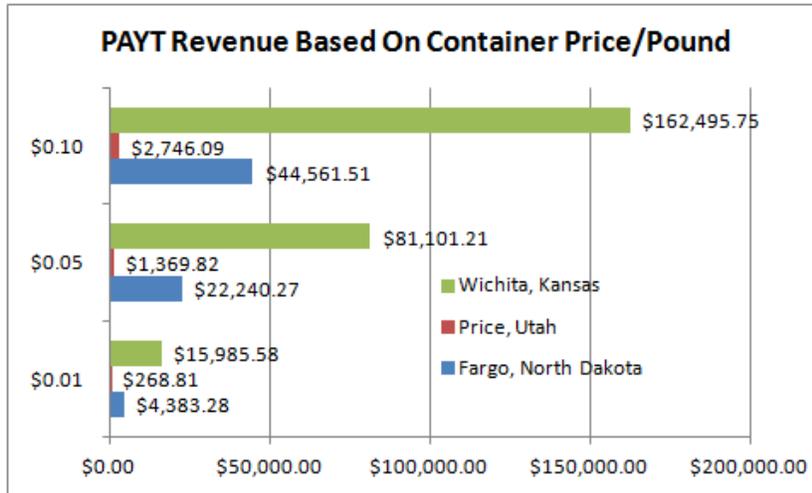
$$\text{Total revenue}_{\text{PAYT}} = (1.125 \text{ lbs} * x - \$0.28/365) * 3047 \text{ people}_{\text{Recycle}} + (4.5 \text{ lbs} * x - \$0.28/365) * (8,402 - 3,047 \text{ people}_{\text{Don't recycle}})$$

Wichita, Kansas

$$\text{Total revenue}_{\text{PAYT}} = (1.125 \text{ lbs} * x - \$0.28/365) * 27,486 \text{ people}_{\text{Recycle}} + (4.5 \text{ lbs} * x - \$0.28/365) * (382,368 - 27,486 \text{ people}_{\text{Don't recycle}})$$

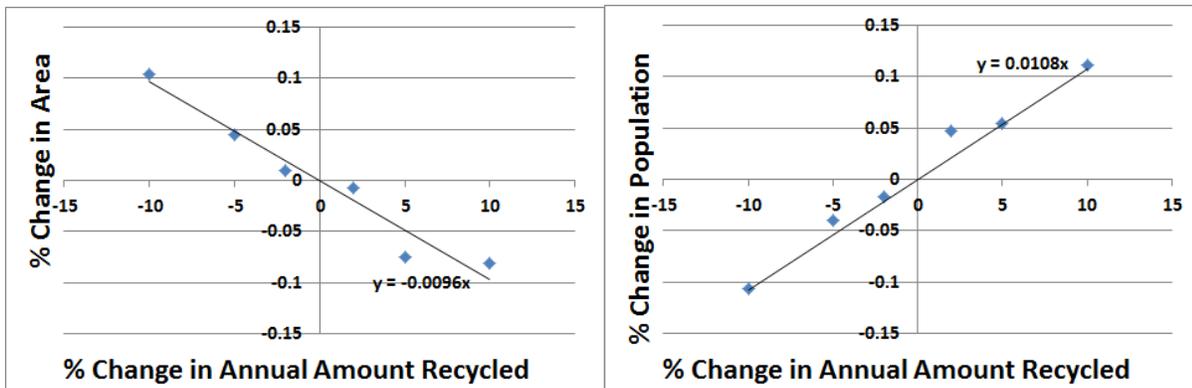
The following table provides the total revenue generated by a PAYT program if the container price per pound were \$0.01, \$0.05, or \$0.10.

Container Price/Pound	Fargo, North Dakota	Price, Utah	Wichita, Kansas
\$0.01	\$4,383.28	\$268.81	\$15,985.58
\$0.05	\$22,240.27	\$1,369.82	\$81,101.21
\$0.10	\$44,561.51	\$2,746.09	\$162,495.75



Sensitivity Analysis

We tested the sensitivity of our simulation of the annual amount recycled in a city using a drop-off recycling program. We changed population and area by +/- 2%, 5%, and 10% and examined the resulting change in annual amount recycled. For simplicity, we only examined the changes for one of our sample cities: Fargo, North Dakota.



The annual amount recycled responds approximately linearly to both area and population. The response is not precisely linear because the randomness used in the simulation to determine whether each household recycles introduces some variation between different runs of the simulation. Since the slopes are small, a slight error in the initial parameters would not significantly change the simulation's output.

Model 2: Curbside Collection*Assumptions*

- 1 Each city has only one recycling processing plant, located at the geographic center, as we found that one large-scale processing center is more than enough to cover one city's recycling needs.
- 2 Recycling collection comes biweekly.

Approach

We subdivided the city into zones for which one garbage truck was responsible. Each truck is responsible for driving to its zone, collecting all the recyclable waste it can, and delivering it to the central processing center, which then sorts and processes the recyclable waste.

Recyclable Waste Collected and Cost to City

We divide the cost to the city into three parts: the cost of gasoline, the wages of the truck drivers, and the price of truck upkeep. The cost is as follows:

$$\text{Cost} = (\text{Price of diesel fuel in dollars/gallon}) * \text{distance} / (\text{Truck miles/gallon}) + (\text{Num houses}) / (\text{Houses/hour}) * (\text{Driver wage/hour}) + \text{Truck_Upkeep}$$

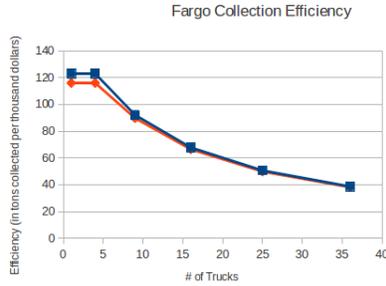
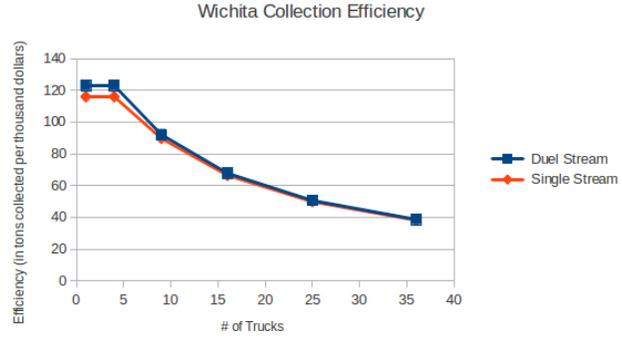
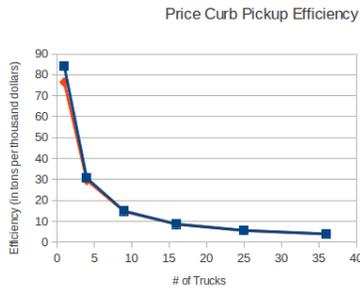
The number of houses visited per hour varies depending on whether a single stream or dual stream collection method is used; for single stream, 171 households are visited per hour, while for dual stream, 130 households are visited per hour [31]. The mileage of a truck is 5 mpg, with a cost of \$4.02 per gallon. The average wage of a truck driver is \$16 dollars/hour [19].

Thus, the formulas for single stream and dual stream collection costs are as follows:

$$\text{Single stream cost} = (\$4.02 \text{ dollars/gallon}) * \text{distance} / (5 \text{ miles/gallon}) + (\text{Num houses}) / (171 \text{ houses/hour}) * (\$16 \text{ wage/hour})$$

$$\text{Dual stream cost} = (\$4.02 \text{ dollars/gallon}) * \text{distance} / (5 \text{ miles/gallon}) + (\text{Num houses}) / (130 \text{ houses/hour}) * (\$16 \text{ wage/hour})$$

We assume that a truck driver can only collect for 7 hours a day: (8 hour work day, minus an hour for lunch and driving). So, a truck driver has a maximum amount of households s/he can visit in a biweekly circuit ($171 * 7 * 10 = 11970$ for single-stream and $130 * 7 * 10 = 9100$ for dual-stream). When a driver is tasked with more houses that s/he can visit, we simply used this ceiling. To demonstrate, the graphs below show efficiency, in terms of tons of recyclable waste collected per thousand dollars, versus the number of trucks in each city.



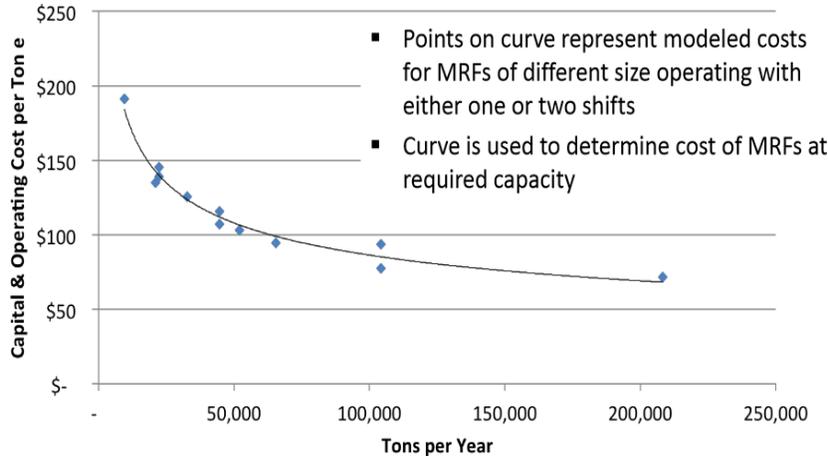
Using the model, we calculated the optimum number of trucks for each city for either dual stream or single stream curbside collection depending on the efficiency of the collection, quantified using the tons of recycled material collected per \$1,000, and the total amount of recycled waste collected. The results for optimum number of trucks are shown below:

City	Single Stream	Dual Stream
Fargo	5	6
Price	1	1
Wichita	13	17

Given the optimum number of collection trucks, the annual cost and tons of recyclable waste collected can then be determined using our computer simulation.

City	Single Stream		Dual Stream	
	Tons of Waste	Collection Cost	Tons of Waste	Collection Cost
Fargo	25292.85	\$205,787.20	25933.39	\$319,383.30
Price	2064.37	\$24,526.54	2064.37	\$27,005.94
Wichita	93947.82	\$713,424.24	88719.32	\$798,496.11

To calculate the revenues and costs generated or incurred from curbside collection, we needed to determine the cost of recycling and sorting at large-scale MRFs. To calculate the net costs per ton of material in processed in a MRF, we used data from Resource Recycling Systems [31] to find operating, capital, and maintenance costs for MRFs of different tonnage capacity. A graphical representation of the processing and operating costs is shown below[31]:



Fargo, North Dakota

Single stream:

Using our model, we calculated that Fargo would generate 25,292.85 tons of recyclables annually using single stream curbside pickup. The operating cost is about \$130 per ton for a dual stream MRF of the same tonnage capacity [31]. However, single stream MRFs have greater processing costs in the range of \$10-15 per ton (averaged at \$12.5), because of greater sorting required [4]. Using the revenue generated from selling recovered material, as calculated in the Drop-Off Center section to be \$128.78, the net cost and total cost are:

$$\text{Net cost per ton} = (\$130 + \$12.5) - \$128.78 = \$13.72 \text{ per ton}$$

$$\text{Total cost} = 25,292.85 \text{ tons} * \$1.22 \text{ per ton} + \text{Collection cost} = \$347,017.90 + \$205,787.20 = \$552,805.1$$

Dual stream:

Using our model, we calculated that Fargo would generate 25,933.39 tons of recyclables annually using dual stream. The operating cost is about \$130 per ton. Thus, net cost and total cost are:

$$\text{Net cost per ton} = \$130 - 128.78 = \$1.22 \text{ per ton}$$

$$\text{Total cost} = 25,933.39 \text{ tons} * \$1.22 \text{ per ton} + \text{Collection cost} = \$31,638.74 + \$319,383.30 = \$351,022.04$$

Price, Utah

Single stream:

Using our model, we calculated that Price would generate 2,064.37 tons of recyclables annually using single stream. A mini MRF, with an annual tonnage of 2,649, is sufficient. The operating cost is about \$245.62 per ton for single stream. Thus, net cost and total cost are:

$$\text{Net cost per ton} = \$245.62 - 128.78 = \$116.84$$

$$\begin{aligned} \text{Total cost} &= 2,064.37 \text{ tons} * \$116.84 \text{ per ton} + \text{Collection cost} = \$24,526.54 + \$241,201 \\ &= \$265,727.53 \end{aligned}$$

Dual stream:

Using our model, we calculated that Price would generate 2,064.37 tons of recyclables annually using dual stream. A mini MRF is again sufficient. The operating cost is about \$218.47 per ton for dual stream. Thus, net cost and total cost are:

$$\text{Net cost per ton} = \$218.47 - 128.78 = 89.69$$

$$\begin{aligned} \text{Total cost} &= 2064.37 \text{ tons} * \$89.69 \text{ per ton} + \text{Collection cost} = \$24,526.54 + \$27,005.94 \\ &= \$212,159.29 \end{aligned}$$

Wichita, Kansas

Single stream:

Using our model, we calculated that Wichita would generate 93,947.82 tons of recyclables annually using single stream. Thus, net cost and total cost are:

$$\text{Net cost per ton} = (\$95 + \$12.5) - \$128.78 = -\$21.28$$

$$\begin{aligned} \text{Total cost} &= 93,947.82 \text{ tons} * -\$21.28 \text{ per ton} + \text{Collection cost} = -\$3,173,557 + \\ & \$713,424.24 = -\$1,285,785.61 \end{aligned}$$

As the cost is negative, the city receives a profit.

Dual stream:

Using our model, we calculated that Wichita would generate 88,719.32 tons of recyclables annually using dual stream. Thus, net cost and total cost are:

$$\text{Net cost per ton} = \$95 - \$128.78 = -\$33.78$$

$$\begin{aligned} \text{Total cost} &= 93,947.82 \text{ tons} * -\$33.78 \text{ per ton} + \text{Collection cost} = -\$3,173,557.36 + \\ & \$798,496.11 = -\$2,375,061 \end{aligned}$$

The city again receives a profit.

Pay-As-You-Throw

We can apply the Pay-As-You-Throw revenue formulas calculated in the Drop-Off Centers section:

$$Total\ revenue_{PAYT} = Revenue_{PAYT, Recycle} + Revenue_{PAYT, Don't\ recycle} = ((4.5\ lbs - 3.375\ lbs) * Container\ price/pound - \$0.28/365) * Population_{Recycle} + (4.5\ lbs * Container\ price/pound - \$0.28/365) * Population_{Don't\ recycle}$$

Given that 40% of people to whom curbside recycling is available recycle [16], we calculated the total revenue each city can expect from a Pay-As-You-Throw program alongside curbside recycling. The variable “x” is used to represent the container price per pound, which is up to the city to set.

Fargo, North Dakota

$$Total\ revenue_{PAYT} = (1.125\ lbs * x - \$0.28/365) * (.40 * 105,549\ people_{Recycle}) + (4.5\ lbs * x - \$0.28/365) * (105,549 - .40 * 105,549\ people_{Don't\ recycle})$$

Price, Utah

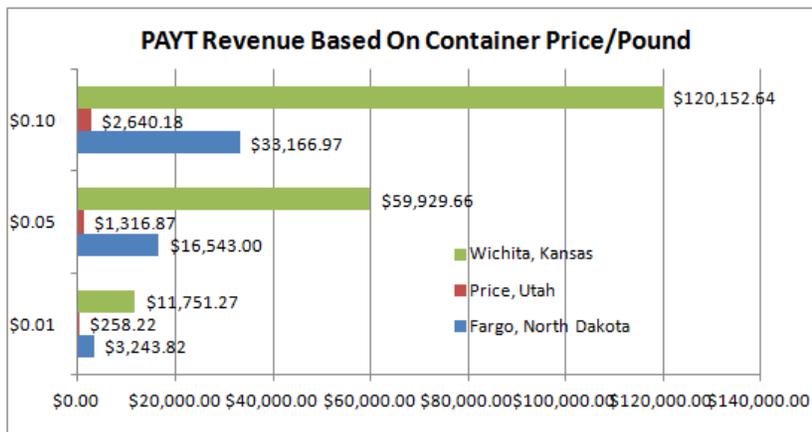
$$Total\ revenue_{PAYT} = (1.125\ lbs * x - \$0.28/365) * (0.40 * 8,402\ people_{Recycle}) + (4.5\ lbs * x - \$0.28/365) * (8,402 - 0.40 * 8,402\ people_{Don't\ recycle})$$

Wichita, Kansas

$$Total\ revenue_{PAYT} = (1.125\ lbs * x - \$0.28/365) * (0.40 * 382,368\ people_{Recycle}) + (4.5\ lbs * x - \$0.28/365) * (382,368 - 0.40 * 382,368\ people_{Don't\ recycle})$$

The following table provides the total revenue generated by a PAYT program if the container price per pound were \$0.01, \$0.05, or \$0.10.

Container Price/Pound	Fargo, North Dakota	Price, Utah	Wichita, Kansas
\$0.01	\$3,243.82	\$258.22	\$11,751.27
\$0.05	\$16,543.00	\$1,316.87	\$59,929.66
\$0.10	\$33,166.97	\$2,640.18	\$120,152.64



3. Testing the Models

To test our models for accuracy, cities with a current drop-off recycling, single-stream curbside collection, or dual-stream curbside collection program can be run through the models. The population, area, and other required attributes of the city will be input into our models, and

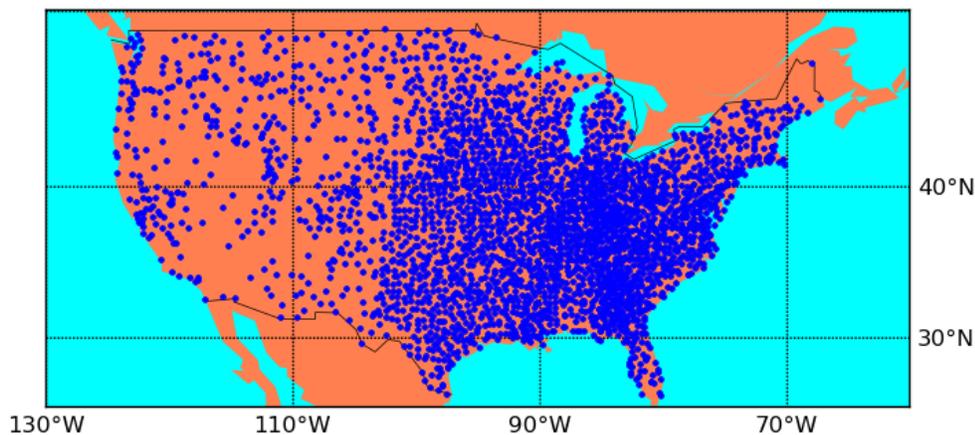
the accuracy of our models would be confirmed if the model results for annual amount recycled and net annual cost to the city are similar to the values in reality.

4. Recommendations

When designing a recycling program, the city should identify markets for recycled materials. The characteristics of the market determine how recyclables should be collected, processed, and eventually sold [32].

To extend our model to the national level, we took US Census data from 2000 which recorded the population density. We then tested each county. On the national level, the EPA should strongly encourage recycling programs for almost every county or region, particularly in denser, less rural regions.

The diagram below marks all the centers of all the counties where at least one of our three proposed recycling programs turns a profit to the community, based on the models were proposed earlier.



In general, across the U.S., very small cities such as Price, Utah will incur losses from a recycling program. A drop-off program cannot be used to full advantage because much of the potential coverage area of one drop-off center lies beyond the city limits. In relatively large, densely populated cities such as Wichita, Kansas, dual-stream curbside collection is generally recommended to bring the highest profits. This is due to low participation in drop-off recycling; on average, only 15.5% of potentially covered households participate. The revenue benefits of a pay-as-you-throw initiative must be balanced against the cost of its unpopularity among citizens. A pay-as-you-throw initiative is generally recommended for small cities such as Price, Utah that seek to adopt a recycling program but incur losses no matter what the program. In these cases, a pay-as-you-throw initiative is recommended to offset losses to the city.

III. CONCLUSION

Effective recycling programs are critical for cities to address the waste accumulation in landfills. Based on our models, we conclude that drop-off centers, curbside collection, and pay-as-you-throw initiatives can all be feasible recycling programs, depending on the population and area of a given city. All the models are resistant to minor changes in the input values and can be applied to any city.

The population growth of the U.S. has a notable effect on the change in the amount of plastic waste discarded in landfills each year. Partly because U.S. population growth has been linear in recent years, we determined that a linear model was most appropriate for predicting the amount of plastic waste discarded. Our linear model predicts that 35.1 million tons of plastic waste will be discarded in 2023, an increase of 13% over 2010.

Using a drop-off program, Fargo, North Dakota, and Wichita, Kansas would both generate profits from the sales of recovered materials. The net profits leave an unpopular PAYT initiative unnecessary. In Price, Utah, however, a drop-off program sustains losses because of the very small size of the city. For this reason, we recommend that Price adopt a PAYT initiative to raise revenues and offset costs of the drop-off program.

Using any curbside program, single-stream or dual-stream, Fargo, North Dakota and Price, Utah incur losses. In both cities, a drop-off program is recommended: in Fargo, because a drop-off generates profits and in Price, a drop-off generates less losses than a curbside collection. In Wichita, Kansas, however, both a single and dual stream curbside collection generate a profit, leaving all three programs feasible. Dual-stream curbside collection is strongly recommended, however, for the highest profits.

Nationally, small cities generally incur losses with any recycling program, as seen in our model results for Price, Utah. Dual-stream curbside collection is generally recommended for large, densely populated cities, who can take advantage of efficiencies of scale. The revenue benefits of a PAYT initiative must be balanced against the cost of its unpopularity among citizens, though it is recommended for small cities to help offset their losses. Recycling has environmental benefits for any city but is especially important for large, densely populated cities, where it has economic as well as environmental benefits.

BIBLIOGRAPHY

- 1 "2010 Facts and Figures Fact Sheet (PDF) - US Environmental ..." 2012. 4 Mar. 2013 <http://www.epa.gov/wastes/nonhaz/municipal/pubs/msw_2010_rev_factsheet.pdf>
- 2 2013. <http://5gyres.org/what_is_the_issue/the_problem/>
- 3 "Municipal Solid Waste | Wastes | US EPA." 2004. 3 Mar. 2013 <<http://www.epa.gov/msw/>>
- 4 "Dual Stream" vs. "Single Stream" Recycling Programs | UMBC ..." 2012. 3 Mar. 2013 <<http://umbcinsights.wordpress.com/2012/08/28/dual-stream-vs-single-stream-recycling-programs/>>
- 5 "Curbside Collection Study - Eureka Recycling." 2010. 3 Mar. 2013 <<http://www.eurekarecycling.org/page.cfm?ContentID=72>>
- 6 "Organizing a Community Recycling Program." 2012. 3 Mar. 2013 <<https://www.bae.ncsu.edu/topic/vermicomposting/pubs/ag473-11-community-recycle.html>>
- 7 "Pay-As-You-Throw Programs| Conservation Tools | US EPA." 2003. 3 Mar. 2013 <<http://www.epa.gov/payt/>>
- 8 "Moving to Fargo, ND| Fargo, ND Moving Companies." 2009. 4 Mar. 2013 <<http://www.upack.com/moving-companies/north-dakota/fargo/>>
- 9 "Explore Utah - Getting Around in Utah - Navigating Utah's Streets ..." 2011. 4 Mar. 2013 <http://www.exploreutah.com/GettingAround/Navigating_Utahs_Streets.shtml>
- 10 Wichita travel guide - Wikitravel." 2005. 4 Mar. 2013 <<http://wikitravel.org/en/Wichita>>

- 11 "A Brief History of Plastic - The Brooklyn Rail." 2008. 4 Mar. 2013
<<http://www.brooklynrail.org/2005/05/express/a-brief-history-of-plastic>>
- 12 "National Intercensal Estimates (2000-2010) - U.S Census Bureau." 2011. 4 Mar. 2013
<<http://www.census.gov/popest/data/intercensal/national/nat2010.html>>
- 13 "Estimating Consumer Willingness to Supply and Willingness to Pay ..." 3 Mar. 2013
<<http://le.uwpress.org/content/88/4/745.refs?related-urls=yes&legid=wple;88/4/745>>
- 14 "Gasoline and Diesel Fuel Update - Energy Information ... - EIA." 2011. 3 Mar. 2013
<<http://www.eia.gov/petroleum/gasdiesel/>>
- 15 "Fuel Economy | National Highway Traffic Safety Administration ..." 2010. 4 Mar. 2013
<<http://www.nhtsa.gov/fuel-economy>>
- 16 "Ohio EPA 2004 Drop-Off Recycling Study." 2010. 3 Mar. 2013
<http://www.epa.ohio.gov/portals/34/document/general/swmd_drop_off_study_report.pdf>
- 17 "Municipal Solid Waste Generation, Recycling, and Disposal in - US ..." 2009. 3 Mar. 2013
<<http://www.epa.gov/osw/nonhaz/municipal/pubs/msw2008rpt.pdf>>
- 18 "Recycling Stats | GreenWaste Recovery." 2008. 3 Mar. 2013
<<http://www.greenwaste.com/recycling-stats>>
- 19 "Enterprise Portal Information - Home." 2006. 4 Mar. 2013
<<http://www.portal.state.pa.us/>>
- 20 "SCRAP TIRE RECYCLING - Tennessee." 2008. 4 Mar. 2013
<<http://www.tn.gov/environment/swm/pdf/TFscaptires.pdf>>
- 21 "Recycled Wood Products - Sonoma Compost." 2007. 4 Mar. 2013
<http://www.sonomacompost.com/recycle_wood.shtml>
- 22 "Plywood Thickness and Weights: Sanded Nominal Thickness 1/4 3 ..." 2008. 4 Mar. 2013
<http://parr.com/PDFs/PG_plywoodthickness.pdf>
- 23 "Recycling and Composting Online." 4 Mar. 2013 <<http://www.recycle.cc/freepapr.htm>>
- 24 "Glass." 2010. 4 Mar. 2013 <<http://www.ndhealth.gov/wm/recycling/Glass.htm>>
- 25 "Daily Scrap Metal Prices - Maxi Waste Limited." 2007. 4 Mar. 2013
<<http://www.maxiwaste.co.uk/scrapdailyprices.php>>
- 26 "Aluminum: How Sustainable is It? | Streamline." 2010. 4 Mar. 2013
<<http://www.streamlinemr.com/articles/aluminum-how-sustainable-is-it>>
- 27 "Textile Recycling FAQs - Trans-Americas Trading Co." 2011. 4 Mar. 2013
<<http://tranclo.com/recycling-faq.asp>>
- 28 "Solid Waste Publications for Sullivan County - Waste Management ..." 2013. 4 Mar. 2013
<<http://waste.uvlsrpc.org/index.php/solid-waste-publications-sullivan-county/>>
- 29 "Georgia Department of Community Affairs." 2007. 4 Mar. 2013
<<http://www.dca.state.ga.us/main/quickmenuListing.asp?mnuitem=PROG>>
- 30 Skumatz, LA. "Measuring Source Reduction: Pay-As-You-Throw Variable Rates as ..." 2000. <<http://www.epa.gov/osw/consERVE/tools/payt/pdf/sera.pdf>>
- 31 "Extending Recycling's Reach with MRF Hub and ... - NC SWANA." 4 Mar. 2013
<<http://www.ncswana.org/files/2012%20Fall%20Presentations/Frey-NC%20SWANA%20MRF%20Optimization%20presentation%20100212.pdf>>
- 32 "Organizing a Community Recycling Program." 2012. 4 Mar. 2013
<<https://www.bae.ncsu.edu/topic/vermicomposting/pubs/ag473-11-community-recycle.html>>

REFERENCES

- [1] S. Chaturapruek, J. Breslau, D. Yazdi, T. Kolokolnikov, and S. McCalla. Crime modeling with Lévy flights. *SIAM Journal on Applied Mathematics*, 73(4):1703–1720, 2013.
- [2] Consortium for Mathematics and Its Applications. 2013 MCM problem A: The Ultimate Brownie Pan. <http://www.comap.com/undergraduate/contests/mcm/contests/2013/problems/>.
- [3] A. Crannell. A Guide to Writing in Mathematics Classes. <http://www.math.wisc.edu/~miller/m221/annalisa.pdf>.
- [4] P. Dawkins. Paul's Online Math Notes: Euler's Method. <http://tutorial.math.lamar.edu/Classes/DE/EulersMethod.aspx>.
- [5] FreeMind. http://freemind.sourceforge.net/wiki/index.php/Main_Page.
- [6] A. Greenleaf, Y. Kurylev, M. Lassas, and G. Uhlmann. Cloaking devices, electromagnetic wormholes, and transformation optics. *SIAM Review*, 51:3–33, 2009.
- [7] D. Hailman and B. Torrents. Keeping dry: Running in the rain. *Mathematics magazine*, 8(24):266–277, 2009.
- [8] Ohio EPA 2004 Drop-Off Recycling Study. 2010. 3 Mar. 2013 http://www.epa.ohio.gov/portals/34/document/general/swmd_drop_off_study_report.pdf.
- [9] M. Stubna, W. McCullough, and G. L. Gray. Euler's Method Tutorial. <http://www.esm.psu.edu/courses/emch12/IntDyn/course-docs/Euler-tutorial/>.
- [10] Y. van Gennip, B. Hunter, R. Ahn, P. Elliott, K. Luh, M. Halvorson, S. Reid, M. Valasik, J. Wo, G. Tita, A. Bertozzi, and P. Brantingham. Community detection using spectral clustering on sparse geosocial data. *SIAM Journal on Applied Mathematics*, 73(1):67–83, 2013.

MOTIVATION AND ACKNOWLEDGMENTS

We came to this project as members of the Moody's Mega Math Challenge's Problem Development Committee, where we have responsibility for soliciting, writing, editing, and vetting potential problems to use in this math modeling scholarship contest. Something that was made clear from the start was that many high school students are unfamiliar with how to do modeling, how to get started, and how to reach a solution. In addition, we have taken part in NSF workshops where the focus is on engaging math education, motivating students to study and pursue careers in STEM, and keeping them in those disciplines when the going gets rough through better understanding and enthusiasm for the subjects.

The support of several individuals and organizations was critical to making the handbook happen: First, Frances G. Laserson, President of The Moody's Foundation, which has funded the Moody's Mega Math Challenge since its inception in 2006. Her commitment to education, specifically in applied math, economics, and finance, has been instrumental in enabling us work on the content contained here. Dr. Peter Turner, Dean of Arts and Sciences at Clarkson University and Vice President for Education and Chair of the Education Committee for the Society for Industrial and Applied Mathematics (SIAM), has supported both the Challenge and SIAM's increased attention to the high school and undergraduate communities of students interested in STEM and, in particular, math modeling. Peter has also been principal investigator, along with Dr. James C. Crowley, executive director of SIAM, on the "Modeling across the Curriculum" workshops funded by the National Science Foundation (NSF) and through which the content of this handbook was also enriched. And finally, Michelle Montgomery, SIAM Director of Marketing and Outreach and Project Director for Moody's Mega Math Challenge, and her staff team at SIAM. Michelle's group organizes the Moody's Mega Math Challenge, participates enthusiastically in the NSF workshops, and seeks to optimize outreach to young people through multiple projects and channels to fulfill one of SIAM's key goals: increasing the pipeline of individuals going into applied mathematics and computational science programs and careers.

#WeDidIt

Karen, Katie and Ben

THIS HANDBOOK WAS MADE POSSIBLE WITH ENTHUSIASM AND FUNDING FROM THE FOLLOWING ORGANIZATIONS:



THE MOODY'S FOUNDATION



Moody's Mega
Math Challenge®



siam®
Society for Industrial and
Applied Mathematics

