

Judge's Commentary

MathWorks Math Modeling Challenge 2021

Kelly Black, Ph.D.

Department of Mathematics, University of Georgia

Introduction

The costs, needs, and infrastructure associated with internet access was the topic for the 2021 Mathworks Math Modeling (M3) Challenge. Teams were provided data for the costs associated with internet access as well as the types of information that is accessed. The students taking part produced outstanding responses and did so under difficult circumstances. Every year the judges are impressed by the determination and dedication of the teams and their coaches. This year the teams managed to persevere with aplomb under circumstances we would never have predicted! More than ever, we are in awe of the teams and their coaches. Thank you all!

In response to the teams' incredible work I will provide an overview of how some teams approached the required tasks in this year's Challenge. The focus of the first three sections is on the three tasks, in order:

- Construct a model to predict the cost of internet access for each of the next 10 years.
- Construct a model to predict the internet needs of a given household and use the model to predict the needs for three given families.
- Construct a method to determine the best locations for stations in a cellular network and demonstrate how it can be used for three different regions.

The final section provides an overview of some of the basic modeling and writing issues the judges observed. Many of these observations have been discussed before, and they are always an issue. This year, however, they seemed to be more acute. I do not know why this is the case but suspect the difficulties associated with coordination and collaboration played a role.

Question One

To address the first question in the Challenge students had to create a model to predict the cost for each unit of bandwidth for each of the next 10 years. Several different approaches were employed, and the majority were guided by the data given to the teams. Some teams constructed a model for the costs associated with internet access for each year, constructed a model for the bandwidth available per year, and then combined the two models by dividing to obtain the cost of access per unit of bandwidth. Other teams used the data to calculate the cost of access per unit of bandwidth and then constructed a model directly from the data.

Once the teams decided which approach to take and which data sets to use, they had to decide on a general form for the model and then determine the parameters using the data. Some teams simply

examined the data to determine a general form. In this situation, though, the question requires the teams to extrapolate into the future, and some basic argument for the general form should be given. For example, a number of teams simply stated they made use of Nielsen's law[1] and included a citation. In doing so they presented a clear justification. Other teams made use of a logistic model for the bandwidth noting that eventually the amount of data that can be transmitted must reach an upper limit due to physical limitations. Once the model is determined it is important to examine the resulting model and discuss the errors (the R^2 value for example), but that alone is not sufficient. Some justification should be given for the choice of the model adopted.

Once a model was developed, a team was expected to also provide some insight into the implications as well as an analysis of the model itself. For example, if an exponential model was developed, then a team should have noted that the model itself implies the costs will eventually get close to zero, which is problematic for long-term approximations of the costs. If a team made use of a logistic model, then a sensitivity analysis of the parameters may have helped identify which approximations for the parameters may show the greatest changes if there is a small change in the values within the data set.

Question Two

To address the second question students had to create a model to determine a given household's internet needs for a year. A demonstration of how to use the model was required as a way to show how the model could be implemented for three different situations. Student teams made use of the data that was provided and assembled the needs for people in an array of different categories. The teams interpreted this in a wide variety of ways. For example, some teams determined the peak bandwidth needs in a given day and used the maximum speed since that determined the infrastructure required to serve the family. Other teams determined the total amount of information required during a day to determine the needs of a family. In either case, a team had to determine what kind of information is accessed for a given demographic group, and for those teams using the maximum speed, they had to also determine when the information was being accessed.

To calculate a result for a given demographic most teams assembled a linear combination of the different kinds of information sources that might be accessed. The coefficients were generally determined by the data. Presenting this kind of information and discussing how the values of the different parameters were chosen was a difficult challenge. Some teams presented their results in many tables while others provided a long narrative of their calculations. It can be difficult to convey this kind of information in either case, and teams that were able to provide a clear narrative that could be easily followed were at a distinct advantage. There is a lot of information to share about the calculations, and one of the basic questions a judge will ask is whether they are able to reproduce a team's results. If the answer to that question is yes, then that tends to create a much more positive impression and instills a greater level of trust in a team's conclusions.

One aspect of the question required students to determine the needs that would meet 90% and 99% of the needs of a family. The first hurdle is that a team must interpret what this means and then clearly convey their interpretation of the requirement. The second hurdle is to determine a distribution of how a given group might access information on the internet. Several teams struggled with this aspect of the question. For example, some teams interpreted this to mean that 90% or 99% of the person's requirements are met. The wording of the question, though, implies that a person's

needs should be fully met except for 10% or 1% of the time. This subtle distinction is important, and a team that provided a response that was consistent with the question made an immediate positive impression.

One surprisingly common approach to address this question was the use of a Monte Carlo simulation. The simulations tended to be performed in one of two different ways. In one of the approaches, people within a group were assumed to have a range of practices and habits with respect to their internet demands. The simulations proceeded by selecting random people within the group and then calculating their needs. The results of such simulations should be consistent with a weighted average value using the probabilities of the different practices, and this is not an efficient way to make use of the computational tools. In the second approach, family groups were chosen at random from different groups, and people within the group tended to have different practices and habits with respect to their internet demands. The simulations were then used to calculate a distribution of needs that could then be used to estimate the 90% and 99% threshold for internet requirements. This represents a more appropriate use of the Monte Carlo method as a way to understand the distribution of needs in a given community. Again, it is important to provide enough information about the method and the associated probabilities that it would be possible for judges to re-implement the scheme on their own, and a judge should not be expected to read the computer code to get the required details.

Question Three

To address the third question students had to determine a model of a wireless, broadband network and then use the model to determine the best placement of the nodes within the network. The teams had to find a way to balance the competing constraints of the capacity required to serve a large number of people trying to access the nodes versus the geographic complexity associated with accessing the nodes over a wide area. Many teams struggled to incorporate both aspects into their model, and many teams focused on only one of the two aspects.

The result is that some of the final recommendations seemed reasonable in a rural area or seemed reasonable in a more densely populated area. A team that was able to construct a model that provided the flexibility to address the needs of both rural and more urban areas tended to make a more positive impression. The basic assumptions that a team adopted tended to result in a wide variety of results, and whether the final recommendations seemed reasonable tended to make a big difference for this question. A judge's initial read of this section and whether the conclusions seemed reasonable tended to make a strong impression when examining this question.

Another important aspect of the third question is that it is closely related to the second question. To address this question the personal habits of the people accessing the network had to first be quantified. Unfortunately, it was not common to see a team able to explicitly provide a close connection between the two questions. Teams that were able to weave a common narrative between the two questions demonstrated that they were able to identify a key aspect of the questions.

Finally, this type of question contains some components that provided a team an opportunity to showcase their understanding of how to use a model once it is established. The question is posed as an optimization problem. The team was expected to achieve some goal and do so in an optimal way.

Such optimization problems are a common task. To address such problems, a team has to explicitly identify two different components. The first component is the set of constraints that must be met. In

this case, the constraints are the requirements associated with providing wireless access to many people spread out over a large area. The second component is the objective function to be minimized or maximized. The objective can be a difficult function to derive and identify. In this case the objective might be to make the costs as small as possible, or the objective might be to maximize the long-term bandwidth available to the customers. It is not uncommon to try to find a way to balance multiple considerations, and in this case it may mean finding a way to make the cost as small as possible while simultaneously providing more than ample internet access. Identifying and clearly discussing the constraints and the objective function, and then figuring out some methodology to determine the best solution, can be a daunting task. A team that was able to discuss a model that met the needs of the community and identified some function to be minimized or maximized clearly demonstrated key insights into the modeling process.

Writing and Modeling

This has been a challenging year for all of us. We recognize that student teams were under a great deal more stress and were likely working in environments that made collaboration difficult. As judges we tried to be mindful and understanding of some of the hurdles that teams faced. We tried to focus on the more basic aspects of modeling as well as how the ideas associated with a model were expressed.

One of the most fundamental practices associated with expressing results is to properly annotate graphs and make consistent use of units. In this year's Challenge, it was more difficult than usual to keep track of the units. In particular, the notion of "bandwidth" had to be clearly defined and annual costs for access had to be calculated. Many teams made use of combinations of the resulting quantities. Additionally, for the second question, many teams made use of linear combinations of different variables. In many cases it was difficult to keep track of the units and going back and forth between different expressions could be a taxing exercise. Clearly labeling units, and more importantly, clearly labeling graphs including appropriate units was even more important in this year's event compared to previous years.

Another important task is to use appropriate computational tools to make repeated calculations. For example, several teams made use of repeated, random sampling to construct a distribution associated with the bandwidth needs of a given population. It was not uncommon for a team to describe their computations by simply stating they "used MATLAB." From the reader's point of view, it can be difficult to determine what this means. At the same time, most readers do not want to comb through computer code to decipher every calculation. It is important to strike a balance and provide some basic overview of an algorithm and to convey a broad sense of the calculations and methods employed to provide an approximation to a given system of equations.

Finally, it seemed to be more common this year to read papers in which the team members' actions were discussed. A report should focus on the model, the analysis of the model, and the results. A discussion of what the team members did or the actions they performed can be a distraction from the fine work that the team performed. The models produced by the team as well as the team's results and conclusions should be the primary focus of the narrative.

Conclusions

Every year we have been impressed and grateful for the hard work of the teams and coaches who make this event an important learning experience. This year more than ever we have been impressed with their steadfast determination. This whole year has been one long series of hurdles, but the teams came forward and put in considerable effort to produce remarkable results. Additionally, the coaches continue to step up and serve and do amazing work despite the extensive demands that have been asked of them throughout the pandemic.

Because of the hard work and dedication of so many people, the teams were able to produce wonderful results. They produced great responses to three tasks associated with the costs, needs, and allocation of resources. A broad overview of some of the responses is provided here, as well as an overview of some general writing and modeling issues. We are always mindful and humbled by the students and coaches that produce these works, and we thank you for all your efforts.

Acknowledgments

I am grateful for the help and insights provided by Kathleen LeBlanc at Society of Industrial and Applied Mathematics. Her guidance and insight are greatly appreciated, and her diligence and aid has greatly improved this document. Thank you!

Bibliography

[1] **Nielsen's Law of Internet Bandwidth**, Doug Dawson, https://www.circleid.com/posts/20191119_nielsens_law_of_internet_bandwidth/#:~:text=One%20of%20the%20more%20interesting,updated%20in%202008%20and%202019. Accessed 12 April 2021.