EXPLORING LAKE ERIE ALGAE GROWTH IN AN ALGEBRA CLASS USING GEOGEBRA

Marie Nabbout-Cheiban

Southern Connecticut State University

Abstract

In the following article, we describe an inquiry-based activity that engages students in an exploration of the spread of toxic algae in Lake Erie. Students collect data showing the algae spread over several days, find best fit curve(s), and make predictions for a follow-up data collection. Students use the data analysis capabilities of GeoGebra to compare their predictions with actual data and adjust their models in an on-going revision process. The software encouraged students to delve into a real-world mathematical problem while making decisions about errors and limitations of their assumptions and potential models.

Keywords: regression analysis, STEM activity, inquiry-based activity

1 INTRODUCTION

In this article, we share informal findings from an activity that was developed and implemented during Project CCSS-STEM, a Teacher Quality Partnership Grant, funded by the Connecticut State Department of Education (CT-SDE) that engages teachers in the creation and use of STEM-based interdisciplinary projects with their students. Small teams of teachers work together with a mentor to design activities based on current and meaningful real-life problems. Together, they create interdisciplinary experiments with data that encourage students to generate potential mathematical models.

1.1 Context for the activity

In August 2014, the presence of toxic algae in Lake Erie made headlines. A group of teachers I was mentoring in Project CCSS-STEM decided to investigate this spread in an Algebra class, exploring best-fit curves to predict future values. The main question that their students explored was as follows: *How can we predict the future area of the Lake Erie algae bloom using measurements of impacted areas from the recent past?* To address such questions, students utilized the data collection and analysis features of GeoGebra in two class sessions over several weeks.

1.2 Overview of student work sessions

1.2.1 Session 1

During the first session, students used GeoGebra to analyze several satellite pictures of the lake taken from the National Oceanic and Atmospheric Administration website (NOAA) (See Figure 1). The pictures show the algae bloom for five different days. Students were asked to gather data and determine the percentage of the infected areas of the lake. Next, students found best fit curves for their data and predicted the spread of the algae for the coming week.

1.2.2 Session 2

The second session took place approximately two weeks later. Students compared their predictions generated in the first session to the real algae growth of the lake and discussed discrepancies in the results and their possible causes (wind effect, human intervention, etc.). They also re-computed their best fit curves to take into account the second round of data and adjusted their predictions. Students engaged in a data collection - model - data collection - revision process, gaining a greater appreciation of the cyclic nature of scientific inquiry. As such, they began to see mathematics as more than simply answering questions on a worksheet. As they explored mathematics within the context of a news story unfolding in their community, they were encouraged to see mathematics study as relevant, meaningful, and on-going.

1.2.3 Student ownership / choice

Before data collection started, students decided whether they wanted to restrict their work to the areas of medium to high pollution or to the larger areas that show more varied pollution levels. To create their lists, students had to detect possible errors and decide which values to include or reject. By providing students freedom in choosing the data to collect and analyze, they felt a sense of ownership with their work that was not typical in the classroom.

1.3 Alignment with the Common Core Mathematical Practices

The activity targets four Common Core Mathematical Practice Standards (MP1, MP2, MP4, MP5). As students analyzed the problem, they looked for “entry points to its solution” and “plan(ned) a pathway rather than jumping into a solution attempt” (MP1). Furthermore, they used their set of data, found best fit curves, and used them to make predictions. To answer the original question, students needed to “decontextualize the problem” and “contextualize the results” (MP2). While working on this activity, students used the mathematics they knew and applied it in a new situation to “solve a problem that arises in society . . . interpreting their mathematical results in the context of the situation . . . reflect(ing) on whether the results made sense, possibly improving the model if it has not served its purpose” (MP4). After each group collected data, they decided on values to be included while rejecting extreme values. As such, they “detect(ed) possible errors by strategically using estimation and other mathematical knowledge” (MP5). GeoGebra was an essential component of the problem solving process for the students. The software enabled them to delve into a mathematical problem while making decisions about errors and limitations of their assumptions and potential models.

2 DETAILED DESCRIPTIONS OF THE ACTIVITY

2.1 Session 1

To initiate the session, video segments, newspaper articles and satellite pictures from 2014 showing the algae growth were presented to the class. Three of six images are shown in Figure 1. Students were asked to collect data from 6 pictures in order to predict the spread of the algae for coming days.

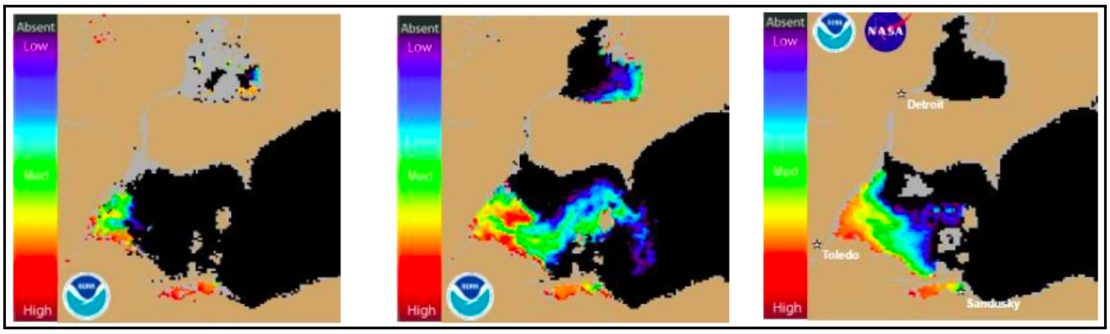


Figure 1. Satellite photos of algae spread from (a) July 31; (b) Aug 3; and (c) Aug 6.

Figure 2 shows an example of the work generated with GeoGebra to calculate the area of the polluted region. Using GeoGebra's polygon tool, students constructed irregular polygons that approximated the shape of the bloom region. Students divided the area of these irregular polygons by the area of the lake (what is shown in figure 2) to calculate the percentage of the area polluted.

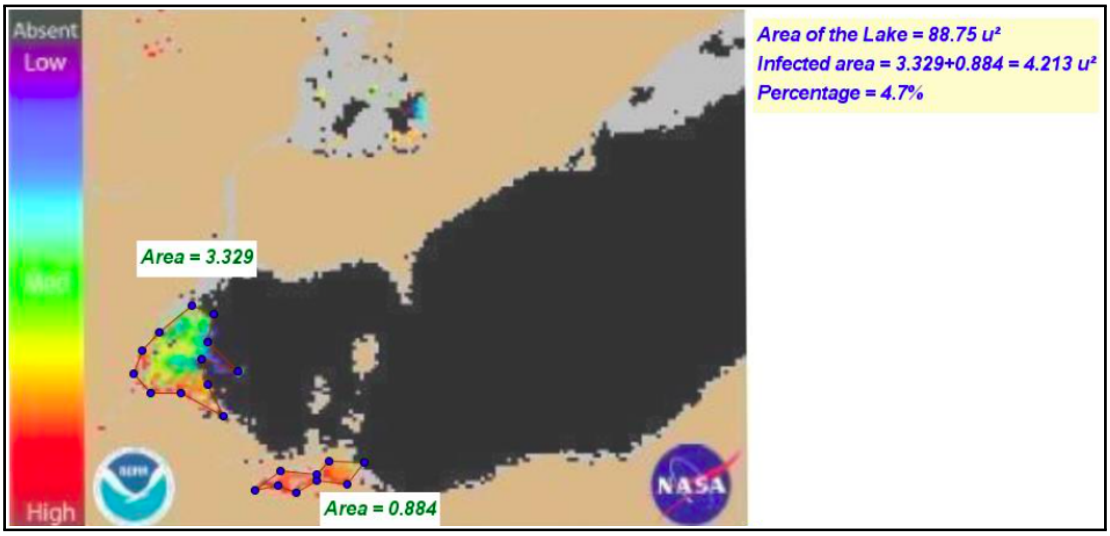


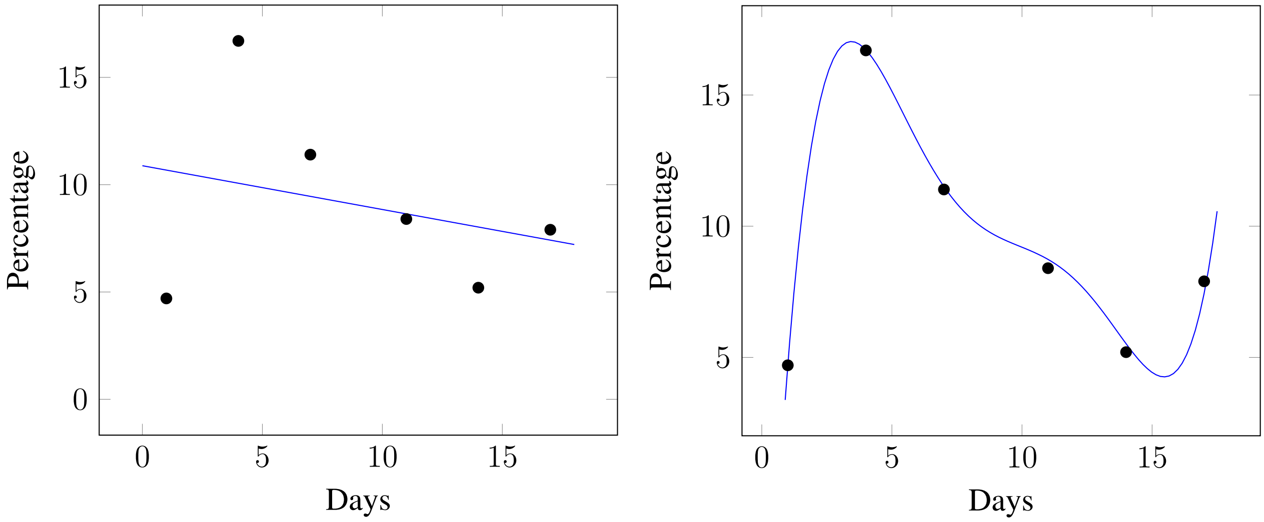
Figure 2. Areas of infected regions calculated with GeoGebra's polygon tool.

Although all students analyzed the same aerial photos, different groups obtained different answers since no two groups surrounded the impacted regions with the exact same polygons and because different groups chose to focus on different areas of algae. After collecting data from images, students constructed two data sets (see Table 1) and found best fit curves for each set.

Date	Day	% (Large Area)	% (Small Area)
July 31	1	4.7	3.6
Aug 3	4	16.7	6.9
Aug 6	7	11.4	6.5
Aug10	11	8.4	5.1
Aug 13	14	5.2	3.9
Aug 16	17	7.9	6.2

Table 1. Toxic algae spread in Lake Erie (July 31 - August 16.)

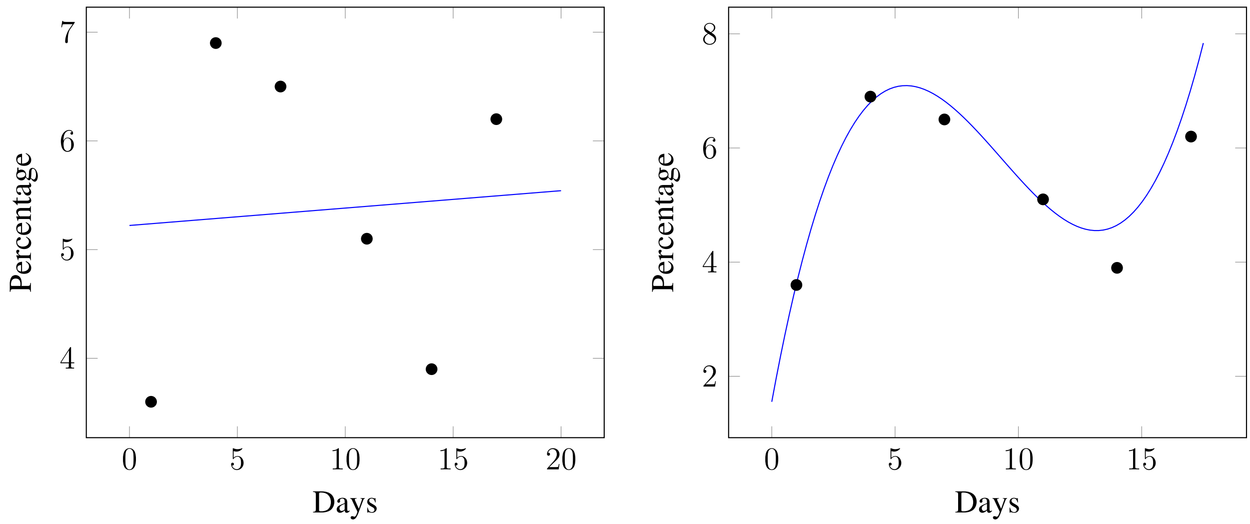
Using GeoGebra, students chose a best fit curve for their data and explained their choice. Figures 3 and 4 show linear and polynomial curves for each set of data.



(a) $y = -0.204x + 10.884$ **(b)** $y = 0.001x^5 - 0.04x^4 + 0.8x^3 - 6.5x^2 + 23.7x - 13.2$

Figure 3. Large Areas: (a) Linear and (b) polynomial models for polluted region.

Using the linear trend curve in Figure 3 (a), the percentage of polluted areas for August 20 and August 26 was predicted to be 6.8% and 5.38%, respectively. Using the polynomial function in Figure 3 (b) as a regression model, one may predict that the lake will be 100% polluted on approximately August 22.

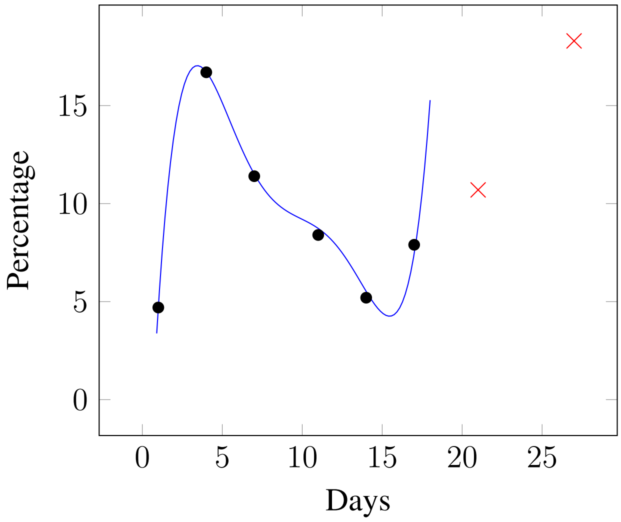


(a) $y = 0.016x + 5.22$ **(b)** $y = 0.011x^3 - 0.307x^2 + 2.363x + 1.553$

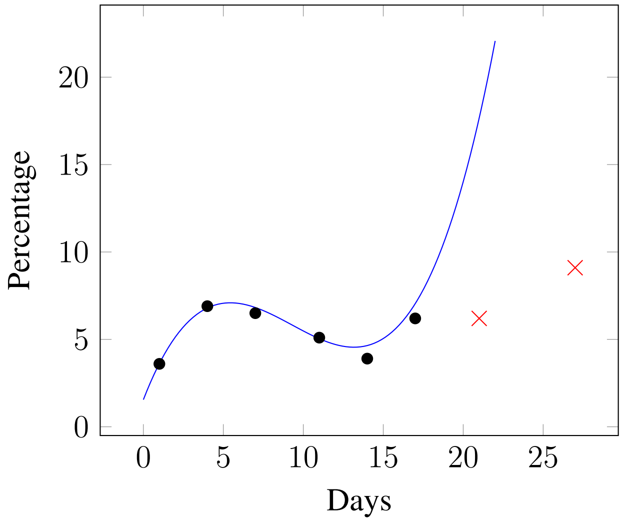
Figure 4. Small Areas: (a) Linear and (b) polynomial models for polluted region.

Using the linear trend line in Figure 4a, the predicted values for percentage areas of pollution for August 20 and August 26 are 5.56% and 5.657%, respectively. Using the polynomial function in Figure 4b as a regression model, one may predict that the entire lake will be highly polluted on approximately Sept 1. Students use the curves in Figures 3 and 4 to predict the spread of algae for the following week as a final activity in Session 1.

It is important to note at this point that although the polynomial curves may pass through every point in the data set and may seem to fit data "perfectly," they are not good models for extrapolation. The main problem lies in the fact that polynomials typically change direction once they leave the dataset under consideration and are only appropriate models on a fixed domain. Allowing students time to explore and understand why they should avoid polynomial curves as predictive models is crucial for strengthening understanding of fit curves. Such discussions took place during Session 2, once students had data that fell outside the reasonable domain of their polynomial models. The red points in Figure 5 show the actual percentage of pollution for August 20 and August 26, 2014.



(a) $y = 0.001x^5 - 0.04x^4 + 0.8x^3 - 6.5x^2 + 23.7x - 13.2$



(b) $y = 0.011x^3 - 0.307x^2 + 2.363x + 1.553$

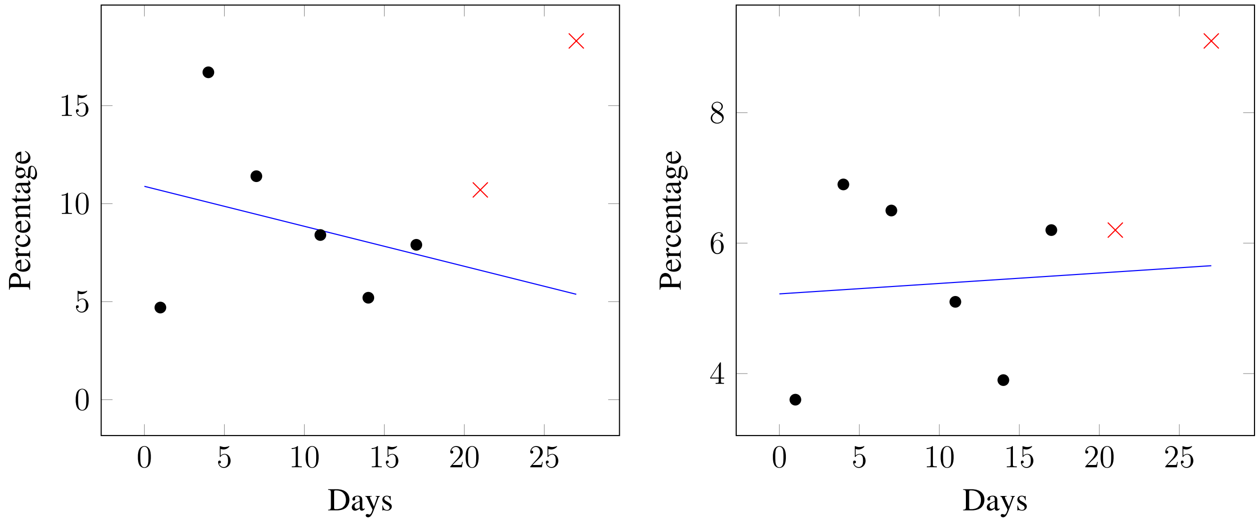
Figure 5. Models generated in Session 1 with actual data shown in red. (a) Large Areas and (b) Small Areas

2.2 Session 2

Session 2 took place approximately two weeks after Session 1. Students were presented with new satellite pictures and percentages of the area of the algae spread. An initial task involved the comparison of actual and predicted values. Students found the best fit line for the large percentage areas had a slightly negative slope, This suggested that, with time, the percentage of the polluted area would decrease. However, the actual data showed the opposite, namely that the percentage of the polluted area increased (see Figure 6a). As Figure 6b shows, the best fit line for the percentage of area with high concentration had a positive slope. Hence students concluded that, with time, the percentage of areas with high pollution would increase slightly. The actual data of the areas with high concentration were somewhat closer to the predictions than the larger areas (see Figure 6).

Next, we asked students to consider factors that affect the algae growth (temperature, sunlight, wind, human impact, etc.) and their relationship to the best fit curves. It is important to note that extrapolation requires a larger set of data. Students included new values in their table of values (see Table 2).

Students recalculated the best fit curves to make new predictions (see Figures 7, 8, 9 and 10). It is important to note that the activity was planned for Algebra 1 class, and students' knowledge of different type of functions and of best fit curves at this level is restricted to best fit lines. Students at more advanced levels are encouraged to employ residual analyses, use inverse operations to linearize data, and use non-linear models.



(a) $y = -0.204x + 10.88$ **(b)** $y = 0.016x + 5.22$

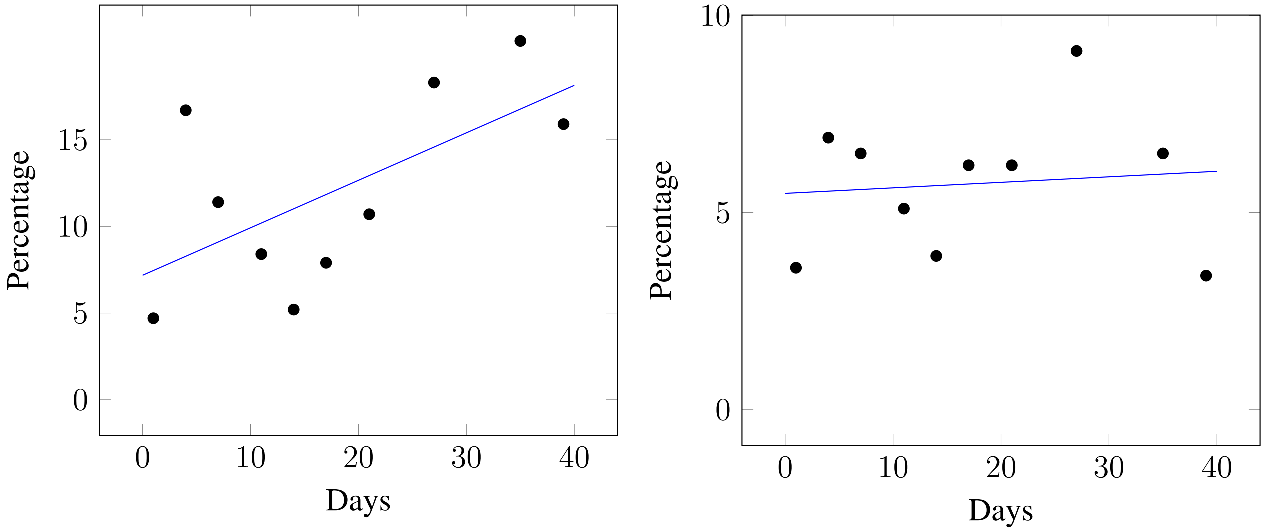
Figure 6. Linear models from Session 1 with actual data shown in red. (a) Large Areas and (b) Small Areas.

Date	Day	% (Large Area)	% (Small Area)
July 31	1	4.7	3.6
Aug 3	4	16.7	6.9
Aug 6	7	11.4	6.5
Aug10	11	8.4	5.1
Aug 13	14	5.2	3.9
Aug 16	17	7.9	6.2
Aug 20	21	10.7	6.2
Aug 26	27	18.3	9.1
Sept 3	35	20.7	6.5
Sept 7	39	15.9	3.4

Table 2. Toxic algae spread in Lake Erie (July 31 - Sept 7.)

For the purpose of this first year algebra lesson, teachers explained why the linear curves do not fit the collected data well, especially for the small areas. Arguably a sine function provides a better fit for the data. Although beyond the scope of this paper, readers are encouraged to verify that the fit is better through an exploration of sum of squared errors for each model.

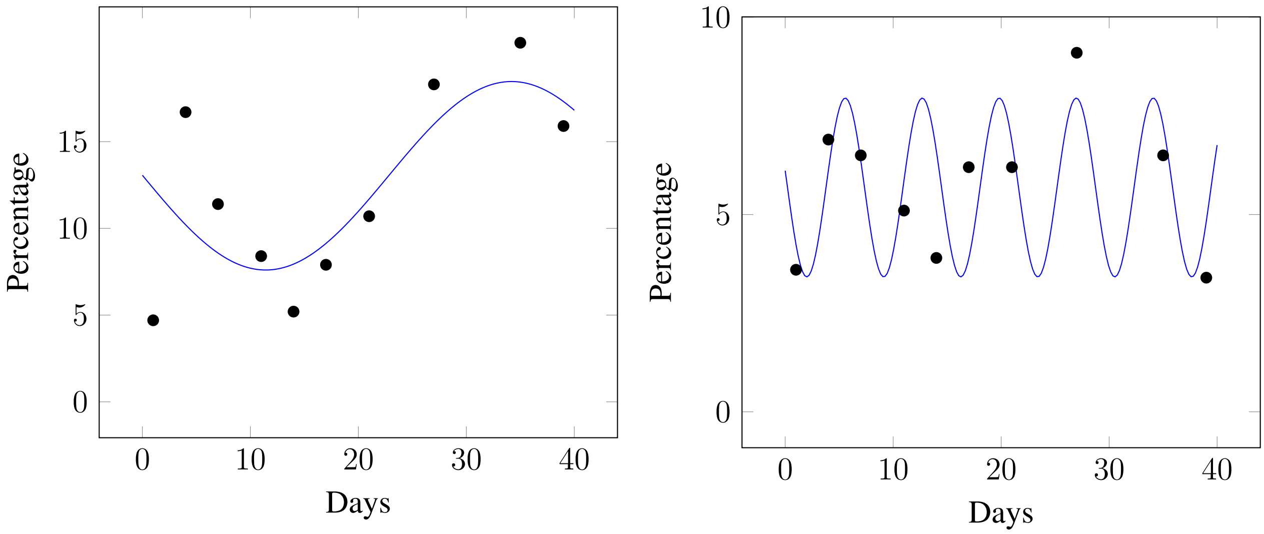
Clearly, better models exist for the data. Exploring several with introductory level students help them, appreciate that real data can be "messy" and its analysis involves careful detective work and some trial and error. Moreover, such experimentation helped students see the effect of the sample size in extrapolation, the importance of validating the one's results before extrapolating, and generalizing. This activity required students' perseverance to solve the problem and allowed them to learn new math that would resolve a word problem.



(a) $y = 0.274x + 7.179$, ($R^2 = 0.38$)

(b) $y = 0.014x + 5.487$, ($R^2 = 0.008$)

Figure 7. Linear models from Session 2 for (a) large areas and (b) small areas.



(a) $y = 13.032 + 5.43\sin(0.138x + 3.136)$, ($R^2 = 0.53$)

(b) $y = 5.68 + 2.26\sin(0.881x + 2.95)$, ($R^2 = 0.496$)

Figure 8. Periodic models for (a) large areas and (b) small areas.

3 Conclusions

In this activity, GeoGebra was used as a data collection tool (to calculate the area on satellite pictures) and as an exploration tool (to predict future values of algae spread). The activity allowed students to use the mathematical concept of area as a "tool" to solve new problems; showing that area can be useful to solve a percentage problem without being the main object of the problem itself. The main problem that teachers faced in this activity was that linear curves did not best fit the collected data. Such problem cannot be predicted ahead of time when dealing with real data, and flexibility in the planning was needed. The description of this activity aims at providing guidelines on how to use a real problem in math classrooms. Many similar activities can be created. One possible source are contexts that are important to students and their environment.

4 Key GeoGebra Command

The following GeoGebra commands were used to generate lines of best fit.

- `FitLine[ <List of Points> ]`
- `FitSin[ <List of Points> ]`

5 ACKNOWLEDGMENTS

I wish to acknowledge Erica Browne and Doreen Mantilia (Daniel Hand High School, Madison, CT) and Michelle Parchesky (Stamford High School, Stamford, CT). Special thanks to the late Betsy Carter for all her helpful comments. Project CCSS-STEM is a Connecticut Teacher Quality Partnership Grant issued by the Connecticut Office of Higher Education. Grant number 130HE1186AA The funding source is the US Department of Education as authorized by PL 107-110, the No Child Left Behind Act of 2001, Title II, Part A, Subpart 3, Improving Teacher Quality State Grant funds.

References

[1] Hohenwarter, M. (2002). *GeoGebra*. Available on-line at `http://www.geogebra.org/cms/en/`

[2] National Governors Association Center for Best Practices and The Council of Chief State School Officers (2010). Common Core State Standards for Mathematics. Washington, DC: Retrieved from URL http://www.corestandards.org/Math/Practice/

[3] National Oceanic and Atmospheric Administration's center for Coastal Ocean Science Bulletin. Retrieved from URL http://www2.nccos.noaa.gov/coast/lakeerie/bulletin/

[4] Wines, Michael (2014, August 7): "Behind Toledo's Water Crisis, a Long-Troubled Lake Erie", New York Times. Retrieved from URL http://nyti.ms/1osw9ND

Marie Nabbout-Cheiban teaches mathematics and mathematics education at the Southern Connecticut State University. She uses technology to create inquiry-based or investigation activities for her students and trains pre-service teachers on teaching using technology. She has designed and conducted several professional development sessions for middle and high school teachers.