The relationship between Big Data and Mathematical Modeling: A discussion in a mathematical education scenario

Rodrigo Dalla Vecchia
rodrigovecchia@gmail.com

Lutheran University of Brazil, Brazil

Abstract. This study discusses aspects of the association between Mathematical Modeling (MM) and Big Data in the scope of mathematical education. We present an example of an activity to discuss two ontological factors that involve MM. The first is linked to the modeling stages. The second involves the idea of pedagogical objectives. The main findings indicate that Big Data may contribute new ways of working with MM in the classroom, helping develop pedagogical objectives associated with the ability to deal with and interpret digital media.

Keywords: Big Data, digital literacy, mathematical education, mathematical modeling

Introduction

Discussing Mathematical Modelling (MM) requires not only understanding aspects linked with the construction and application of specific models but also seeing MM as an ever-changing development and analyzing its components from the ontological standpoint. Despite the consistent research on this topic carried out in recent years, MM is still open to a wide array of interpretations (Dalla Vecchia, 2012; Klüber, 2012; Maltempi & Dalla Vecchia, 2013). This well-known multiplicity of interpretative perspectives, which obstructs the consolidation of a shared view, has become even more complex with the advent of Digital Technologies and Communication and Information Technologies that, for Lévy (1996, pp. 17-18), may lead to a:

"*[…] shift in the ontological center of gravity of the object considered: instead of defining itself mainly based on its actuality (a 'solution'), the entity begins to find its essential consistency in a problematic field*".

Lévy's words (1996) highlight the fact that technologies may influence the fundamental characteristics of the entity, situation, or object being analyzed, changing the way we understand it. In the MM context, this involves different repercussions in the search for solutions to the problem investigated. In a scenario characterized by various technological expressions, Big Data attracts attention due to the possible implications it may have in the efforts to understand MM more fully.

According to IBM (2011), Big Data has to be considered in the context of the treatment of very large databases that often require different resources and methodologies, when compared with standard data. Of the several perspectives and problematics surrounding Big Data, we are interested in the ideas associated with the use of such data as an element of the production of new information about a given phenomenon. In this sense, we understand that Big Data "*[…] is more than a mere question of size; it represents an opportunity to gain insights into new data types and contents, […] and to answer questions that until recently were left outside the scope of Big Data*" (IBM, 2011).

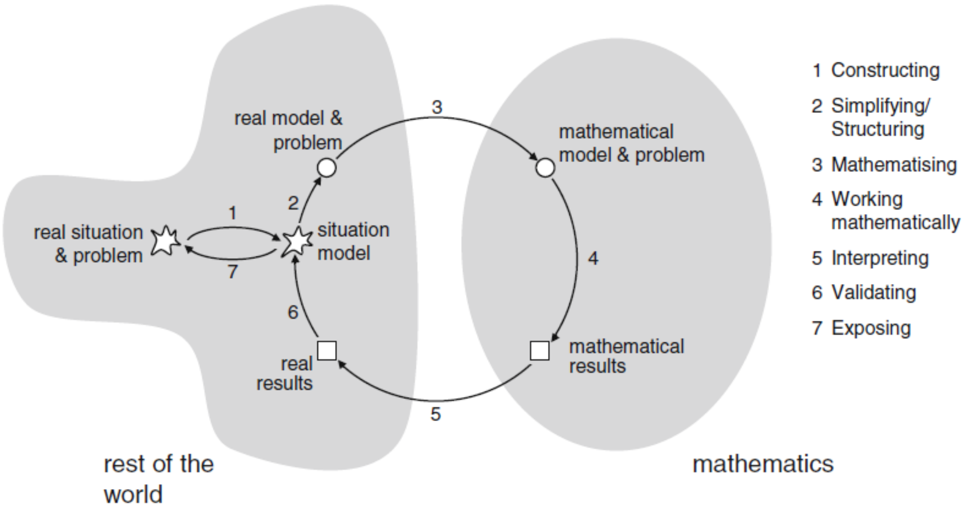

Figure 1. Modelling cycle from a cognitivist perspective

So, MM stands as a remarkable approach to treating massive data volumes. It has several applications, such as organizing and reorganizing data, changing data characteristics, drawing inferences, and recognizing determined types of phenomena. In this article, we see Big Data as a research instrument that may potentially elicit a reconsideration of ontological aspects of MM.

Following this line of thought, the present article looks into some of the most singular aspects of MM when phenomena that originate from Big Data are mathematically investigated. More specifically, we discuss MM processes in terms of pedagogical objectives, searching for a suitable theoretical framework for this purpose. As a means to support these argumentations, we discuss some examples quantitatively (Lincoln & Guba, 1985).

Analyzing how MM processes are understood

A review of the research by authors like Borromeu Ferri & Blum (2010) shows that a MM process is often seen as a sequence of predefined steps (Figure 1). According to these authors, these steps are sequentially taken as soon as the task is assigned. The first step in this series consists of building a model in order to figure out a situation, which is then simplified, structured, and idealized, when associations are made between the situation evaluated and mathematics. From this idealization, the structure is interpreted from the mathematical standpoint and mathematically treated until results are reached, which likewise are of mathematical nature. These results are then interpreted in light of the actual situation and subsequently validated. This cycle ends with the presentation of the results obtained. When no such results are achieved, the cycle is restarted.

With minor differences in such visions, Bassenai (2004), Biembengut & Hein (2007) and Kaiser, Schwartz & Tiedman (2010) offer a contribution to this discussion claiming that MM is a sequential process of advancing steps. However, in a study on the cyberworld, Dalla Vecchia (2012) already observed changes in this process when contrasted with the notion presented by Borromeu Ferri & Blum (2010). In our previous study, the participants used a programming language to build models to move an object on a map. During the model updating process, the participants noticed that the object was not moving along the path marked on the map. In an MM process as defined by Borromeu Ferri & Blum (2010), the model should be reviewed in the attempt to replicate the actual situation. However, what happened was exactly the opposite: the students decided to keep the model as it was, changing the path marked on the map. In other words, it was the reference system that was adapted to fit the actual model. As a result, when the reference system is the experienced reality, the model is refuted. Conversely, in the cyberworld, this reference system may be refuted, though the main idea in the given situation is preserved in particular aspects.

As subtle as this difference may seem, it indicates that the space created by technologies may elicit changes in the way the MM process is understood. As in Dalla Vecchia (2012), we believe that Big Data is a prospective tool to rethink the MM process in classroom scenarios.

Below we present an example to illustrate this difference. Let us suppose that we are interested in working with second-degree polynomial functions. The general expression of these functions is

$f: R \rightarrow R$ where $f(x) = ax^2 + bx + c$; $a, b, c \in R$; and $a \neq 0$.

This function affords to model several phenomena, especially using approximation techniques (as in the least squares method studied in Calculus, for instance). Despite the various models for this function, Big Data tools such as Google Correlate allow developing MM activities that nevertheless behave differently, considering Borromeu Ferri & Blum (2010).

For Dos Santos & Lemes (2014), Google Correlate detects search behaviors that follow the standards that best fit a set of predefined time and place data series. This means that the tool uses web search data to identify the searches that follow similar standards in a destination data series. The results are made available for consultation or download as a CVS file in the Google Correlate website.

In order to better understand Google Correlate and the role it may play in second-degree polynomial functions, let us consider the function

$f: R \rightarrow R$, where $y = f(x) = x^2 - 8x + 16$.

In the classroom, a table listing independent variable values (x) and the corresponding dependent variable values (y) is used to explain the graph of this category of function. Next, the tabulated values are represented as dots on a Cartesian plane, that is, the function graph (Figure 2).

x	y
0	16
1	9
2	4
3	1
4	0
5	1
6	4
7	9
8	16

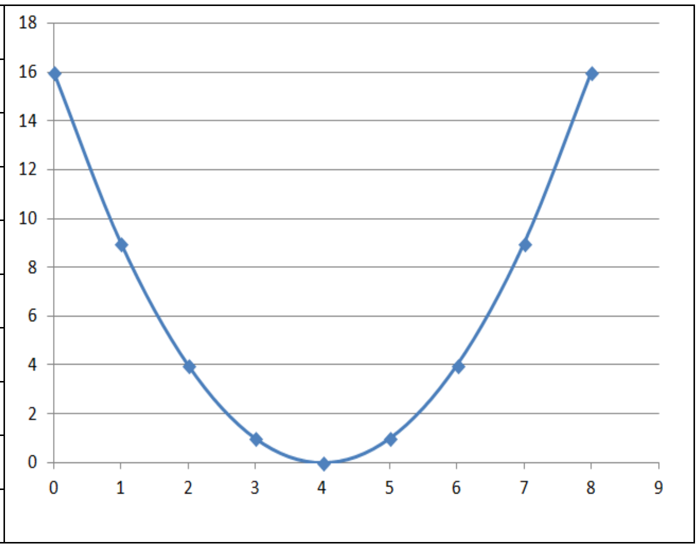

Figure 2. Construction of a polynomial function graph

As said above, there is a close relationship between this category of function and MM. Nevertheless, for Bassanezi (2004) and Biembengut & Hein (2007), associating a phenomenon we want to study to a specific mathematical content sometimes is no simple task in a classroom. If we follow the steps suggested by these authors, which are similar to those proposed by Borromeu Ferri & Blum (2010), we should begin with a real situation, then isolate it and adapt it to mathematics, create one or more models that may replicate the situation we wish to model, solve the mathematical problem obtained, interpret the solutions in light of the situation investigated, and, finally, compare these solutions with the real situation.

However, Google Correlate also affords to associate the model built in the example with real internet search situations and to find the best correlation with the function considered. This may be done in two ways. With a stronger educational potential, the first is based on the use of Google Correlate's tool 'Search by Drawing', which enables to draw a graph similar to the one prepared in the example and that provides the best correlation with the drawing. Interestingly, even when all students use the same graph as a reference system in a classroom, different correlations may be obtained due to the particular nature of each drawing used. Figure 3 shows the graph we prepared to illustrate the example given above.

In the scope of the present article, we chose to look for correlations associated only with searches carried out in Brazil, the author's country of origin. The resulting drawing is shown in Figure 4.

The second way to work using Google Correlate involves producing time-based datasheets in software like Excel, for example, and exporting the data for online analysis. Google Correlate will look for the best correlation between data on the spreadsheet and searches on specific subjects on the web.

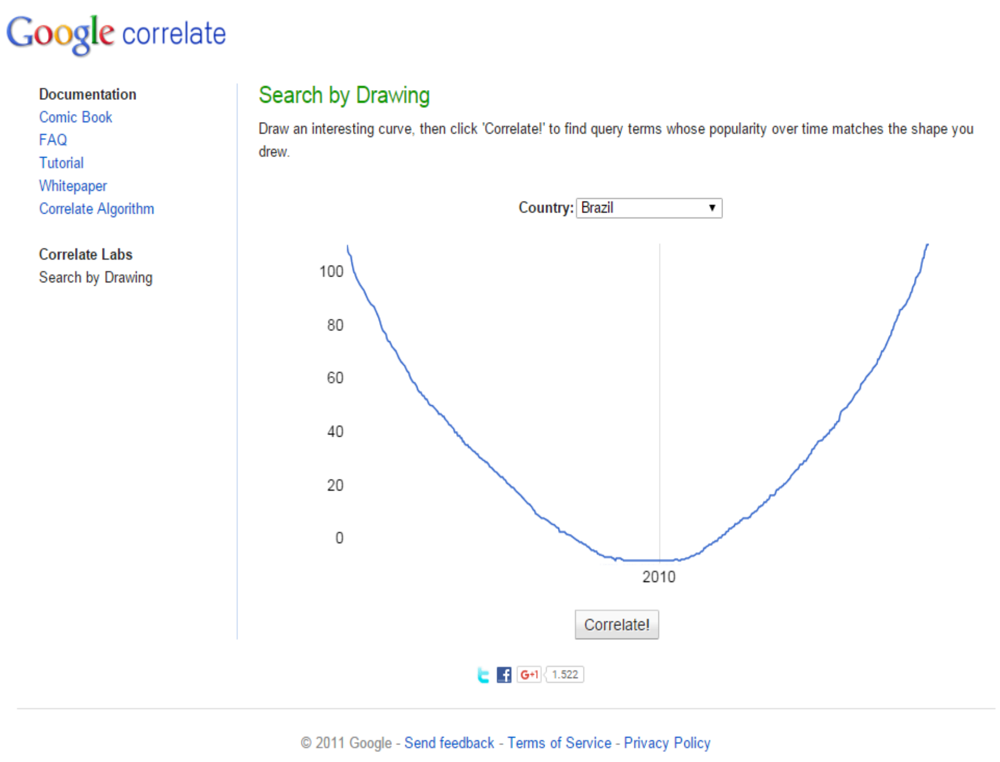

Figure 3. Drawing similar to the graph of the polynomial function described in the example

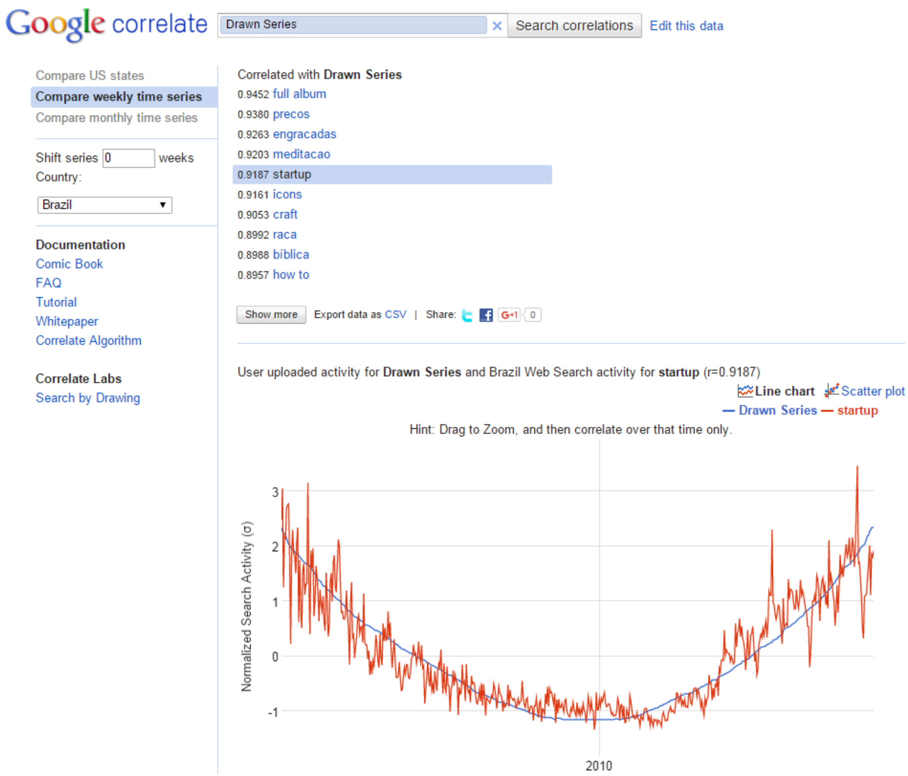

Figure 4. Google Correlate result

Independently of the way we chose, in the present study, we highlight the possibility to identify actual behavior associations that are similar to the model previously built. As reported by Dalla Vechia (2012), the way MM is conducted goes through a 'rupture', or 'inversion' process. While Borromeu Ferri & Blum (2010) start with a real problem or situation and then proceed to develop a model, in the example described above the starting point is the actual development of the model, after which we try to find a real situation that may be demonstrated by the model. More specifically, it is possible to find not one but several fitting situations with a correlation with the model proposed. Therefore, in addition to rethinking the whole MM process, we understand that working with Big Data in Google Correlate also affords to contextualize a given mathematics subject.

Yet, from the educational perspective, where exactly lies the importance of activities like those we discuss in the present study? What benefits can we obtain? In order to find answers to these questions we have first to discuss some ideas about *pedagogical objectives* in the MM context, indicating the pathway to fully develop activities.

Mathematical modeling and pedagogical objectives

Considering the current Mathematical Education scenarios, the present study understands that MM is "*[...] a dynamic and pedagogical process of constructing models supported by interrelated mathematical ideas whose aim is to solve problems in any dimension embedded in reality*" (Dalla Vecchia, 2012, p. 218). This notion of MM presupposes the existence of four fundamental aspects that compose the model construction process, namely, the pedagogical objective, models and language, the problem, and reality. Using a metaphor, in that study the author claims that the multiple characteristics of each aspect become intertwined, influencing the MM process like a stone that creates waves when thrown into a lake (Figure 5).

Figure 5. MM seen as an evolving stream mediated by the multiplicity afforded by the model

Figure 5 shows that the waves do not form an isolated field; rather, it is the fields that actually affect one another, creating streams. When assessed from this perspective, MM may be seen as a static process, since any change caused in it may decisively influence the pathway to a solution to the problem. Regarding the four aspects presented, we underscore the importance of *pedagogical objectives* as '[…] the set of targets or purposes to be met during the development of a proposal side by side with students aiming towards the educational process (Dalla Vecchia, 2012, p. 71).

This view takes inspiration from the notion of *educational process*, understood by authors like Iturra (1994, p. 2), as being "*[…] the means through which those who already consolidated the hows and whys of their own historical experience attempt to rescue the younger ones out of the inconsistency of their knowledge of what is perceived, though not made explicit; […] [trying to] enclose the younger ones in cultural taxonomies*".

For Iturra (1994), the educational process also may be seen from two distinct perspectives. The first is associated with the notion of teaching what we have been producing, especially what we already know. The second is based on the understanding of how knowledge is produced, with special emphasis on this construction process so that learners may think, position themselves accordingly, and develop solutions to problems faced in different scenarios. This means that, in the first perspective, '[…] the educational process is a reiteration of what we already know, while the second perspective is based on a thought structure that may explain the alternatives to solve the questions faced in life' (Iturra, 1994, p. 2). In the scope of the concepts adopted in the present study, we see a pedagogical objective as being more closely related to the second perspective to interpret the educational process.

In the context of MM, the importance of pedagogical objectives lies mainly in their roles in directing the whole process. This is the notion supported by authors like Barbosa & Santos (2007, p. 2), when they state that '[…] different purposes imply differences in the ways to develop and carry out MM activities'. In addition, this orientation does not necessarily mean a predefined path; rather, it implies a compass point for the multiple directions the MM process may take.

The view supported by Barbosa (2001) is a good example of this orientation. For the author, MM provides the means to "read reality critically", based mostly on the social-critical notion advocated by Skovsmose (2000, 2006, 2007). However, Bassaneze (2004) and Biembengut & Hein (2007) have a different point of view, claiming that MM is a *teaching methodology* whose main objectives concern the construction of mathematical knowledge. In turn, for Malheiros (2008), MM is a means to *construct democratic thought*, in that MM should be used according to a differentiated notion of curriculum, in which problems are discussed by students. Therefore, critical readings of reality, teaching methodologies, and democratic media are examples of the different pedagogical objectives associated with MM. However, in the correlation between MM and Big Data, what are the pedagogical objectives that could be covered? Is the target a mere contextualization of content?

Inspired by the ideas supported by Dos Santos & Lemes (2014), we further consider the example of MM proposed in the present study in view of educational objectives that agree with current and future technological challenges. In their investigations, Dos Santos & Lemes (2014) proposed the use of Google Correlate as a tool to prepare students (i) to the scientific challenges presented by Big Data in the real world, and (ii) to a better comprehension of the notions of phenomenon, observation, measurement, physics laws, theory, and causality. The activities proposed included the search for correlations between terms chosen by the students themselves, which nevertheless had some relationship with Physics Teaching. The main objective was to find plausible scientific explanations (causations) for the correlations observed. The results show that students engaged themselves in activities, and demonstrate that they understand the differences between correlation and causation.

Following these ideas, instead of ending the MM process at the moment we recognize a correlation associated with the model proposed, we proceed with the search for causal relationships in the behavior of the phenomenon observed, further exploiting the example proposed in this article. Based on the data in Figure 5 showing associations with the words 'full album', 'prices', 'funny', 'meditation', and 'startup', we frame the questions: Why was there a concern about prices in 2004? Why did this concern drop to a low in 2010 and then again to another minimum in 2015? Is there a new concern about this aspect? Is it associated with the economic scenario in the country? What are its origins? Is there an association between the correlated words?

In this sense, we understand that the MM process proceeds towards a different direction when compared with what is proposed by Borromeu Ferri & Blum (2010), when it does not advance towards data validation, given that these data have been shown to correlate with the model. The movement takes place towards understanding the phenomenon modeled based on cause-and-effect relationships and scientific explanations. More specifically, we believe that this kind of educational position may be linked to what Jenkins (2006) calls *digital literacy*, which is understood as the ability to deal with and interpret digital media. The authors discuss this idea claiming that the current social and historical scenario, which is immersed in the technological world, creates new needs and requires skills that have to be addressed by the education environment. In this sense, they maintain that children and teenagers are engaged in a process of building skills and competencies by interacting with media, and that these abilities are not taken into account by the education environment. This set of skills include:

- Play: the ability to experiment with the medium and to use it to solve problems.
- Performance: the ability to change, aiming to improvising and discovering new things.
- Simulation: the ability to interpret and build dynamic models based on the real world.
- Appropriation: the ability to experiment and to reorganize digital contents and with the aim of using them.
- Multitasking: the ability to analyze the medium so as to perceive important surrounding details and so use them.
- Distributed cognition: the ability to interact significantly with resources that enable personal growth.

- Collective intelligence: the ability in which the student reaches conclusions about a subject on a personal level, and compares it against the notions supported by his/her peers based on a critical analysis in search for an objective in common.
- Judgment: the ability to evaluate the reliability and the credibility of different information sources, since these abound in the digital environment.
- Transmedia navigation: the ability to search, summarize, and impart information.
- Negotiation: the ability to move across different communities, telling apart and respecting different perspectives while following alternative norms.

We understand that, by proposing an activity based on Google Correlate in MM in Mathematical Education, we may set pedagogical objectives that go beyond the teaching of Mathematics, seeing it as a means to meet objectives associated with *digital literacy*. These objectives may be met developing the abilities described by Jenkins (2006) that, in our opinion, are directly linked with the proposal of activity presented in this study.

Final Considerations

In this article, we discussed the possible associations between MM and Big Data in Mathematical Education. We presented a conjectural but realistic example to understand the MM process and proposed the search for correlations that were similar to the graph of a second degree polynomial function.

As simple as it is, the process proposed critically analyzes the classic MM notions, as in Borromeu Ferri & Blum (2010), Bassanezi (2004) and Biembengut & Hein (2007). While, for these authors, the point of departure is a real problem or situation and only then a model is sought, in the example introduced here we begin with the model and then look for a real situation that may be modeled using the same model. In particular, we may find not one but several situations with good correlation with the model proposed. Therefore, we understand that using Big Data and Google Correlate may be convenient resources not only to contextualize a mathematical topic but also to rethink the MM process.

In addition, we took the effort further, searching for causal comprehensions of the correlations listed by Google Correlate. This stage of this investigation was inspired by the works of Dos Santos & Lemes (2014), who addressed these associations when teaching Physics, reaching promising results. We understand that this effort to understand this phenomenon may be directly linked with pedagogical objectives aligned with digital literacy. So, we believe that the idea of an association between MM and Big Data may open new horizons in the educational process, which relate to digital literacy and whose aim is to develop the required skills to deal with current and future changes triggered by Digital Technologies and Communication and Information Technologies.

Our current and future studies, in addition to the aspects discussed in the present paper, address considerations of ontological aspects linked with the reality of the cyberworld, the relationship between the model and this reality in the context of Big Data, and the emergence of new information on large data volumes.

References

Barbosa, J. C. (2001). *Modelagem Matemática*: Concepções e Experiências de Futuros Professores. São Paulo: UNESP, 2001. Tese (Doutorado em Educação Matemática), Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista, Rio Claro.

Bassanezi, R. C. (2004) *Ensino-aprendizagem com Modelagem Matemática*. São Paulo: Contexto (2nd edition).
Biembengut, M. S. & Hein, N. (2007). *Modelagem Matemática no Ensino*. São Paulo: Contexto.
Borromeu Ferri, R., & Blum, W. (2010). Insights into teachers' unconscious behaviour in modeling contexts. In R. Lesh, P. Galbraith, C. R. Haines, A. Hurford (eds.). *Modeling Students' Mathematical Modeling Competences* (pp. 423-432). New York: Springer.
Dalla Vecchia, R., & Maltempi, M. V. (2009) Ensaio sobre a modelagem matemática e o virtual. In *Anais do XIII Encontro Brasileiro de Estudantes de Pós-Graduação em Educação Matemática* (pp. 1-15). Goiânia.
Dalla Vecchia, R., & Maltempi, M. V. (2010) Tecnologias digitais e percepção da realidade: contribuições para a modelagem matemática. *Anais do X Encontro Nacional de Educação Matemática* (pp. 1-10). Bahia.
Dalla Vecchia, R. (2012). *A Modelagem Matemática e a realidade no mundo cibernético.* Rio Claro: UNESP, 2012. Tese (Doutorado em Educação Matemática), Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista. Rio Claro.
Dalla Vecchia, R., & Maltemp M. V. (2012). Modelagem Matemática e Tecnologias de Informação e Comunicação: a realidade do mundo cibernético como um vetor de virtualização. *Bolema*, 26(43).
Dos Santos, R. P., & Lemes, I. L. (2014). Aprender-com-Big-Data no Ensino de Ciências**.** *Acta Scientiae*, 16(4), 178-198.
IBM. (2011). *What is big data?* Armonk, NY: IBM. Retrieved 10 May 2015, from http://www-01.ibm.com/software/data/bigdata.
Jenkins, H. (2006). *Confronting the Challenges of Participatory Culture: Media education for the 21st Century*. The MacArthur Foundation, Chicago, 2006. Retrieved 19 August 2015, from https://www.macfound.org/media/article_pdfs/JENKINS_WHITE_PAPER.PDF.
Kaiser, G., Schwarz, B., & Tiedemann, S. (2010) Future Teachers' professional knowledge on modeling. In R. Lesh, P. Galbraith, C. R. Haines & A. Hurford (eds.)**.** *Modeling Students' Mathematical Modeling Competences* (pp. 433-444). New York: U.S.A., Springer.
Klüber, T. E. (2012). *Uma metacompreensão da modelagem matemática na educação matemática*. Florianópolis: UFSC - Universidade Federal de Santa Catarina. Retrieved 3 August 2015, from https://repositorio.ufsc.br/handle/123456789/96465.
Lévy, P. (1996). *O que é o virtual*. São Paulo: Editora 34.
Lincoln, Y., & Guba, E. (1985). *Naturalistic Inquiry*. California: Sage Publications.
Maltempi, M. V., & Dalla Vecchia, R. D. (2013). About mathematical modeling in the reality of the cybernetic world. In B. Ubuz, Ç. Haser & M. A. Mariotti (eds.) *Proceedings of the Eight Congress of the European Society for Research in Mathematics Education* (pp. 109701106). Antalya: Middle East Technical University, Ankara.
Malheiros, A. P. S. (2008). *Educação Matemática Online*: a elaboração de projetos de Modelagem. São Paulo: UNESP. Tese (Doutorado em Educação Matemática), Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista, Rio Claro.
Skovsmose, O. (2000). Cenários para Investigação. *Bolema*, *13*(14), 66-91.
Skovsmose, O. (2006). *Educação Matemática Crítica***:** a questão da democracia. 3 ed. Campinas, SP: Papirus.
Skovsmose, O. (2007). *Educação Crítica*: incerteza, matemática, responsabilidade. São Paulo: Cortez.

To cite this article: Dalla Vecchia, R. (2015). The relationship between Big Data and Mathematical Modeling: A discussion in a mathematical education scenario. *Themes in Science and Technology Education*, 8(2), 95-103.

URL: http://earthlab.uoi.gr/theste