Reference Models: Using Models to Turn Data into Evidence

Teru Miyake*†

Reference models of the earth's interior play an important role in the acquisition of knowledge about the earth's interior and the earth as a whole. Such models are used as a sort of standard reference against which data are compared. I argue that the use of reference models merits more attention than it has gotten so far in the literature on models, for it is an example of a method of doing science that has a long and significant history, and a study of reference models could increase our understanding of this methodology.

1. Introduction. Reference models of the earth's interior play an important role in the acquisition of knowledge about the earth's interior and the earth as a whole. Such models are used as a sort of standard reference against which data are compared. Deviations between the observations one would expect if the reference model were an accurate representation of the earth and actual observations are used to make inferences about the earth's interior. Perhaps the most widely used such model in geophysics, the Preliminary Reference Earth Model, or PREM (Dziewonski and Anderson 1981), was completed in 1981, and it was utilized for the construction of many other models through the end of the twentieth century (Ritzwoller and Lavely 1995).[1]

*To contact the author, please write to: Nanyang Technological University, 14 Nanyang Drive, School of HSS 06-05, Singapore 637332; e-mail: tmiyake@ntu.edu.sg.

†Much of the research for this article was done under Startup Grant No. 4080821 at Nanyang Technological University in Singapore. I would like to thank Lina Jansson for helpful discussion. Some of the ideas in this article date back to research done at Stanford University, where I received the advice and encouragement of George Smith, Michael Friedman, Tom Ryckman, and Helen Longino. Finally, I would like to give special thanks to Pat Suppes, whose contributions are immeasurable.

1. See Smith (2007) for an account of the history of seismology leading up to the construction of PREM.

Philosophy of Science, 82 (December 2015) pp. 822–832. 0031-8248/2015/8205-0008$10.00

There is a recent, growing literature focusing on the use of models in science (e.g., Morgan and Morrison 1999; Wimsatt 2007; Weisberg 2013). The use of models in a manner similar to the way in which reference models are used in geophysics is described by Wimsatt (2007), but he mentions these uses merely in passing in his discussion of neutral models in biology. Weisberg (2013) has a much more comprehensive and systematic account of models and their uses in science, but he does not specifically mention a use of models in the manner I will describe in this article. I will argue that the use of reference models merits more attention than it has gotten so far in the literature on models, for it is an example of a method of doing science that has a long and significant history, one that has recently been described by Smith (2002, 2014) and Harper (2011) as "turning data into evidence," and a study of reference models could increase our understanding of this methodology.

The aim of this article is to contribute to the literature on models by first locating reference models relative to the general taxonomy of models described by Weisberg and comparing them to the use of neutral models in biology as described by Wimsatt. I then examine some possible desiderata for the construction of reference models, and I end the article with some considerations about the connection between reference models and "turning data into evidence."

2. Models and Idealization. I start with Weisberg's picture of models because it is the most ambitious recent attempt to give a comprehensive account of models and their use in science, and it appears likely itself to become a standard reference on models for philosophers of science. From the standpoint of Weisberg's picture, earth reference models would best be construed as target-directed models that utilize Galilean idealization. Target-directed models are models for which the modeler has a specific target in mind. For earth reference models, the target is clearly the interior of the earth. In Weisberg's picture of models, there are three different ways in which models can be idealized: Galilean idealization, minimalist idealization, and multiple-models idealization. Galilean idealization involves the simplification of models with the aim of making them more mathematically tractable. Minimalist idealization involves the construction of models that include only difference-making factors that are necessary for a phenomenon, with the aim of constructing an explanation of a given phenomenon. Multiple-models idealization involves building multiple incompatible models of a single phenomenon, usually in the study of highly complex phenomena.

As we shall see, earth reference models involve Galilean idealization, so I want to examine this notion in more depth. Weisberg's discussion of Galilean idealization (2013, 99) depends heavily on the description given in McMullin (1985). Typically, there is some phenomenon of interest, but it is too complicated to model faithfully, so an initial simplified model is created. Then, this simplified model is used to improve our understanding of the phenomenon, and the simplified model is gradually made more realistic in a process that McMullin calls "de-idealization." Weisberg takes the whole purpose of Galilean idealization to be to deal with intractability, and thus "advances in computational power and mathematical techniques should lead the Galilean idealizer to de-idealize" (2013, 99).

Weisberg does not give very detailed examples of this process of de-idealization, but McMullin does.[2] The most detailed example he gives is the Bohr model of the hydrogen atom (McMullin 1985, 260–61). The Bohr model, in which the electron is in a circular orbit around the proton, could be used to predict the energy levels of the electron, which could then be compared to spectroscopic observations of hydrogen. More specifically, a theoretical value for the Rydberg constant could be calculated, which could then be compared to empirical measurements of this constant. McMullin says that at least three idealizations were being made here: the neutron is at rest, the orbit of the electron is circular, and relativistic effects are left out. Later on, successive corrections were made to the model that, McMullin claims, resulted in a closer fit between the model and reality. McMullin describes this process as one where the model "serves as the basis for a continuing research program," one in which the model starts off as a tractable model that has significant departures from reality, and this model is gradually filled in with more and more details.

Here I want to ask exactly how the initial model serves as a basis for this research program. There are two significantly different ways in which it could do that. The first way is for the model simply to provide a sort of initial sketch upon which further and further new details are added. These details might come about through new observations, or through the development of new mathematical or computational techniques that overcome the intractability problems that led to the development of the initial simplified model, allowing such details to be filled in, where previously they could not. The second way is for the model itself to be used directly to produce the new observations from which the further details can be added. I call the first kind of process *passive* de-idealization, while I call the second kind *active* de-idealization. We will see that earth reference models are used for active de-idealization.

2. McMullin makes distinctions of his own regarding idealization, such as that between formal and material idealization. The Bohr model of the atom is given as an example of formal idealization. McMullin's distinctions might well cross-cut Weisberg's distinctions, and I do not want to complicate the picture here, so I will refrain from any discussion of McMullin's distinctions.

Exactly how does active de-idealization work? Although Wimsatt (2007) does not use my terminology, he describes an example of active de-idealization. One of the major points that Wimsatt makes is that false models can be used in many different ways to learn true facts about complicated systems. He gives a list of 12 ways in which false models can be used to search for better models. I want to focus here on the first five such functions he gives for false models (Wimsatt 2007, 104):

1. An oversimplified model may act as a starting point in a series of models of increasing complexity and realism.
2. A known incorrect but otherwise suggestive model may undercut the too-ready acceptance of a preferred hypothesis by suggesting new alternative lines for the explanation of the phenomena.
3. An incorrect model may suggest new predictive tests or new refinements of an established model, or highlight specific features of it as particularly important.
4. An incomplete model may be used as a template, which captures larger or otherwise more obvious effects that can then be "factored out" to detect phenomena that would otherwise be masked or be too small to be seen.
5. A model that is incomplete may be used as a template for estimating the magnitudes of parameters that are not included in the model.

The first function is, of course, mentioned by both Weisberg and McMullin. It is a statement of the idea of Galilean idealization and the process of gradual de-idealization. In functions 2 and 3, a false model is used as a heuristic—it suggests "new alternative lines for the explanation of phenomena," or "new predictive tests or new refinements." I want to focus particularly on functions 4 and 5. When used for these functions, Wimsatt says that the false model is used as a "template" that is used either to factor out larger effects, in order to capture effects that are too small to be seen, or for estimating parameters that are not themselves included in the model.

The discussion in Wimsatt (2007) involves a detailed study of the linear linkage model developed by Thomas Hunt Morgan in the early twentieth century. Wimsatt gives several examples of cases where deviations from the predictions of the linear linkage model were used to postulate causal factors that were not being taken into account in the model. This use of the model would fall under function 4 (Wimsatt 2007, 106–11). He also discusses a case where deviations from the predictions of another model, the Haldane mapping function, were used to estimate the value of a parameter that is not contained in the model itself (Wimsatt 2007, 120).

I want to emphasize again that functions 4 and 5 for models as described by Wimsatt are cases of active, not passive, de-idealization. The false model is used directly to produce evidence that can then be used to extract information about the system or phenomenon of interest. It is not being used merely as a heuristic—rather, the model itself is used to produce the observations. Wimsatt provides a very good example of these uses of false models, but one might get the impression that the way in which models are used here is relatively rare in science. This impression, however, is mistaken—there are at least some sciences where this is the primary way in which progress is made. Most of our knowledge of the interior of the earth, for example, is the result of the application of this method.

Perhaps one of the reasons that this method has not gotten the attention it deserves is that it raises some rather difficult issues with regard to justification. There is, first of all, a circularity worry. Suppose I create an initial model, and then I study the deviations from this model. These deviations are then taken to be evidence for, say, causal factors that must be taken into account in the model. I then add these further causal factors and improve the model. Perhaps I then investigate further deviations from the predictions of this new model and try to make further improvements to the model. If, however, the wrong initial model was used, then the deviations might not reflect any real causal factors after all—they might turn out to have been illusory, in which case the research program would have been going down an illusory "garden path."[3] So one thing you would want to be careful about is that if a false model is being used to create new observations, the model ought to be false in the right way—the deviations from the model ought to be ones that actually will tell us true things about the system, or at least point us in the right direction. You would then expect that there might be some norms for models if they are being used in this way. I discuss such possible norms below. I now turn to a discussion of earth reference models—models that I believe are used in the way I have described.

3. Earth Reference Models. The earth reference models that I have mentioned in this article are idealized models of the mechanical properties of the earth's interior. If the interior of the earth is taken to be elastically isotropic, then the mechanical properties of each point in the earth's interior can be characterized by three variables: density, and two parameters that express the elastic properties of the medium, usually incompressibility and rigidity. If the earth is taken to be spherically symmetric, that is, mechanical properties of the earth are taken to depend only on the distance from the center of the earth, then the mechanical properties of the entire earth can be represented completely in terms of three functions of radius. For such an

3. This is George Smith's term. See Smith (2002).

idealized earth, expected travel times for various types of seismic waves can be calculated. In the 1930s, spherically symmetric models of the mechanical properties of the earth's interior were determined by constructing idealized earth models and comparing expected travel times for such models with actual travel times of seismic waves. There was a remarkable agreement between the earth models produced by the two main groups working on earth models at the time, one involving Harold Jeffreys and Keith Bullen, and the other involving Beno Gutenberg and Charles Richter (Bullen 1975). The methods used here were hypothetico-deductive—that is, the models were postulated as hypotheses about the earth's internal structure, and they were compared directly against observations of travel times.

In the 1960s, the fortuitous confluence of digital computing technology with a couple of the largest earthquakes ever recorded made possible the recording of normal modes of oscillation of the earth. Normal mode frequencies can be calculated for an idealized, spherically symmetric earth, and models that incorporate normal mode frequency observations were built starting in the 1960s. Further, the advancements in computing allowed geophysicists to develop Monte Carlo methods in which earth models were generated randomly by computer and tested against observations (Press 1968). Some of the models that agreed with observation were significantly different from any other model that had been postulated at the time, and these studies led to worries about the possibility of radically different models being consistent with observations. Work by the geophysicists George Backus and Freeman Gilbert (Backus and Gilbert 1967, 1968, 1970), which tried to address this nonuniqueness problem, showed that limits could be put on the degree of nonuniqueness of earth models, but only under the assumption that the functions relating earth structure to observations of normal mode frequencies were linear, an assumption that was known to be false.

According to the geophysicist Keith Bullen (1974), a committee was set up in 1971 for the construction of a "Standard Earth Model." The reason given for the construction of this new model was that a large amount of new data had been collected since the Jeffreys–Bullen and Gutenberg–Richter models had been constructed, and individual geophysicists had been incorporating these new data in different ways. This had led to a "great untidiness in the presentation of numerical seismological results." In the mid-1970s, several teams of geophysicists began to develop earth models with the goal of coming up with a standard reference model. In 1981, this process culminated with the development of PREM, which is still being used to this day, although there are now several other alternative models that are used as reference models as well.

Earth reference models, such as PREM, are used in many ways, but what is most distinctive about their use from an epistemological point of view is that they are best thought of in terms of functions 4 and 5 in the taxonomy of functions of idealized models described by Wimsatt. They are used, that is, for detecting phenomena that would otherwise be masked or be too small to be seen, or for estimating magnitudes of parameters that are not included in the model.

These two uses can be seen quite clearly in the way in which PREM has been used for the construction of three-dimensional models of the interior of the earth, that is, models that are no longer simply spherically symmetric, but express the mechanical properties of the earth's interior in terms of three spatial variables. Most of these models are based on observations of travel times of seismic waves. They are not, however, constructed by simply building a model and comparing its predictions with actual travel times of seismic waves. The observations used are usually travel time residuals—that is, the deviations from the travel times predicted by a reference model such as PREM. The three-dimensional model constructed is then a linear perturbation of a one-dimensional reference model, such as PREM (Ritzwoller and Lavely 1995). Thus, the deviations between the observations predicted by PREM and actual observations are being used to identify three-dimensional features of the earth that are not in PREM itself and to measure parameters that represent mechanical properties of such additional features.

4. Possible Norms for Reference Models. I now want to think about possible norms that might govern the use of reference models, keeping in mind Wimsatt's functions 4 and 5: detecting phenomena that would otherwise be masked or be too small to be seen, or for estimating magnitudes of parameters that are not included in the model. Reference models are being used to produce new observations through an analysis of the deviations between actual observations and predicted observations of the model. These observations are then used to eventually arrive at a better picture of the earth's interior. In order to be useful in this process, the models are idealized—that is, they are false, but they must be false in the right way. What is "false in the right way," though? There are, I suggest, two primary norms. First, the models must be simple in such a way that they can be utilized easily in this process of producing further observations. Second, the models must somehow reflect the physical situation, in such a way that deviations between what they predict and actual observations actually have some kind of physical significance.

Here is an example of how the first norm played into the development of PREM. In the mid-1970s, there were several teams of geophysicists working on different earth models toward the development of the standard reference model. One such model was a "parametrically simple earth model" (Dziewonski, Hales, and Lapwood 1975). This spherically symmetric model represented the mechanical properties of the interior of the earth in terms of a piecewise continuous function, where most of the pieces were low-order polynomials. There is, of course, no reason to think that the mechanical properties of the earth are truly distributed in accordance with low-order polynomial functions. There are, however, advantages in this kind of representation. For example, the "travel times of body waves and their derivatives would always vary smoothly as a function of distance on a particular branch of a travel time curve" (Dziewonski et al. 1975, 12). As I mentioned above, one-dimensional reference models are often used for the construction of three-dimensional earth models using travel time residuals as observations. A model in which the predicted travel times varied smoothly as a function of distance would be easier to compute residuals for. This would be useful not only for the construction of three-dimensional models but also for other investigations that require the use of travel time residuals, such as the location of seismic sources. Ultimately, the representation of large sections of the interior of the earth in terms of low-order polynomials was adopted into PREM (Dziewonski and Anderson 1981) as well.

The other norm is, I think, more complicated. Reference models must be false, but they must be false in a physically meaningful way. Often, what this means is that reference models will not be the best-fit model empirically. One of the geophysicists involved in the construction of PREM, Adam Dziewonski, discusses this consideration in a later article that considers the possibility of constructing a new reference model: "A reference model, in a modern sense, is one which satisfies more than just one class of seismological or geophysical observations—like, for instance, travel times of body waves. It should constitute a common basis of reference for all the different studies concerning the earth. . . . This strategy seeks a model which has to be physically meaningful—as opposed to an empirical one, which could achieve good results at reproducing a narrow range of observations rather than explaining them" (Morelli and Dziewonski 1993, 179). What is meant here by "physically meaningful" is that deviations between what the model predicts and what actual observations show give us useful information about the interior of the earth.

Exactly what "physically meaningful" means could depend on the specific ways in which the reference model is being used. For example, if the reference model is used in the construction of three-dimensional models of the earth, it would ideally correspond to the lowest-order term in a spherical harmonic expansion of the normal modes of the earth. Deviations from such a model would contain information about higher-order modes that would be indicative of finer three-dimensional structure. However, if a reference model is going to be used for many different purposes, a more general notion of "physically meaningful" might have to be used. This is a complicated matter, on which further work needs to be done. Here, however, I would like to point out the connections between the use of reference models and some recent work on scientific methodology.

5. Turning Data into Evidence. In this final section, I want to discuss connections between the way in which reference models are used in geophysics and some recent work on scientific inference by George Smith (2002, 2014) and William Harper (2011).[4] Both Smith and Harper have done extensive work on Newton, and they both emphasize the role of Newton's fourth rule for philosophizing in Newton's methodology: "In experimental philosophy, propositions gathered from phenomena by induction should be considered either exactly or very nearly true notwithstanding any contrary hypotheses, until yet other phenomena make such propositions either more exact or liable to exceptions" (Newton 1999, 796).

The fourth rule of reasoning says two things: that we should rule out hypotheses in favor of propositions that are gathered from phenomena, and that we should provisionally take such propositions to be either exactly or very nearly true. Smith (2002, 2014) argues that Newton's methodology involves taking such propositions gathered from phenomena to be provisionally true so that deviations between what you would expect the phenomena to be like, given that the propositions are true, and what the phenomena are actually observed to be like can be found. These deviations are then taken to be new phenomena that require explanation. Both Smith and Harper refer to this process as "turning data into evidence." They both reject a simple hypothetico-deductive picture where there is a hypothesis, and this hypothesis is supported (or rejected) by data. Instead, certain propositions are needed in order to extract phenomena from raw data—to "turn data into evidence."

The parallel with the use of reference models is straightforward. Reference models play the role of propositions gathered from phenomena. Expected observations for these reference models are calculated as if the reference models were true, and then deviations between these expected observations and actual observations are either taken to be indicative of further causal factors or used to try to measure further parameters that are not captured in these models. Reference models are being used, in other words, to turn data into evidence. "Turning data into evidence" is another term for what I have been calling active de-idealization.

If this is, indeed, an accurate picture of a significant way in which science is done, then it might be useful to think about the norms that govern this methodology. If one of the aims of building models—or, more generally, theorizing—is to enable active de-idealization, then we might expect the norms that are required here to be different from those that would govern

4. George Smith has, himself, written on earth models (Smith 2007), including PREM, although his focus is on the period in geophysics before the construction of PREM, and not on the uses of PREM and other reference models.

a more standard picture where models or theories are constructed without active de-idealization in mind. For example, there might be a norm for simplicity that is driven less by notions about the connection between simplicity and truth, or by simple tractability considerations, than by the fact that models that are simpler in certain ways can more easily be put to use in producing further observations. There might also be a fairly complicated norm for "physical meaningfulness"—one that requires a model to yield deviations that would tell us something about a system or phenomenon of interest.

Now, one notable difference between Newton and earth modelers is that earth modelers already have fairly good ideas about what "physical meaningfulness" amounts to when building earth models, although they might not, by any means, have a complete picture. On the other hand, the whole difficulty for Newton was coming up with a background theory that would allow him to differentiate between what is "physically meaningful" and what is not. Thus, one might think that what I have to say here about reference models does not easily apply to the case of Newton. On the contrary, however, I believe that a detailed examination of the use of reference models could, itself, be a useful reference against which to compare the difficulties faced by Newton and others in various important episodes in the history of science.

REFERENCES

Backus, George, and Freeman Gilbert. 1967. "Numerical Applications of a Formalism for Geophysical Inverse Problems." *Geophysical Journal of the Royal Astronomical Society* 13:247–76.

———. 1968. "The Resolving Power of Gross Earth Data." *Geophysical Journal of the Royal Astronomical Society* 16:169–205.

———. 1970. "Uniqueness in the Inversion of Inaccurate Gross Earth Data." *Philosophical Transactions of the Royal Society of London* A 266 (1173): 123–92.

Bullen, Keith. 1974. "Introductory Remarks on Standard Earth Model." *Physics of the Earth and Planetary Interiors* 9:1–3.

———. 1975. *The Earth's Density*. London: Chapman & Hall.

Dziewonski, Adam, and Don Anderson. 1981. "Preliminary Reference Earth Model." *Physics of the Earth and Planetary Interiors* 25:297–356.

Dziewonski, Adam, Anton Hales, and E. R. Lapwood. 1975. "Parametrically Simple Earth Models Consistent with Geophysical Data." *Physics of the Earth and Planetary Interiors* 10:12–48.

Harper, William. 2011. *Isaac Newton's Scientific Method: Turning Data into Evidence about Gravity and Cosmology*. Oxford: Oxford University Press.

McMullin, Ernan. 1985. "Galilean Idealization." *Studies in History and Philosophy of Science* 16 (3): 247–73.

Morelli, Andrea, and Adam Dziewonski. 1993. "Body Wave Traveltimes and a Spherically Symmetric P- and S-Wave Velocity Model." *Geophysical Journal International* 112:178–94.

Morgan, Mary, and Margaret Morrison. 1999. *Models as Mediators*. Cambridge: Cambridge University Press.

Newton, Isaac. 1999. *Mathematical Principles of Natural Philosophy*. Trans. I. B. Cohen and Anne Whitman. Berkeley: University of California Press.

Press, Frank. 1968. "Density Distribution in Earth." *Science* 160 (3833): 1218–21.

Ritzwoller, Michael, and Eugene Lavely. 1995. "Three-Dimensional Seismic Models of the Earth's Mantle." *Reviews of Geophysics* 33 (1): 1–66.

Smith, George E. 2002. “The Methodology of the *Principia*.” In *The Cambridge Companion to Newton*, ed. I. Bernard Cohen and George E. Smith. Cambridge: Cambridge University Press.

———. 2007. “Gaining Access: Using Seismology to Probe the Earth’s Insides.” http://www.stanford.edu/dept/cisst/events0506.html.

———. 2014. “Closing the Loop: Testing Newtonian Gravity, Then and Now.” In *Newton and Empiricism*, ed. Zvi Biener and Eric Schliesser. Oxford: Oxford University Press.

Weisberg, Michael. 2013. *Simulation and Similarity*. Oxford: Oxford University Press.

Wimsatt, William. 2007. *Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, MA: Harvard University Press.

https://doi.org/10.1086/683322 Published online by Cambridge University Press