Studies in History and Philosophy of Modern Physics 41 (2010) 218–232

Contents lists available at ScienceDirect

Studies in History and Philosophy of Modern Physics

journal homepage: www.elsevier.com/locate/shpsb

Lumping, testing, tuning: The invention of an artificial chemistry in atmospheric transport modeling

Matthias Heymann

Department of Science Studies, University of Aarhus, C. F. Møllers Allé 8, 8000 Aarhus C, Denmark

ARTICLE INFO

Article history:
Received 2 February 2010
Received in revised form
5 July 2010
Accepted 15 July 2010

Keywords:
Atmospheric chemistry
Computer simulation
Computer model
Model building
Air pollution
Ozone formation
Atmospheric transport

ABSTRACT

Since the late 1950s computer simulation has been used to investigate the transport of pollutants in the atmosphere. About 20 years later also the chemical transformation of atmospheric pollutants was included in computer models of photochemical smog formation. Due to limited knowledge of atmospheric chemistry and due to limited computer capacity, chemical processes in the atmosphere were modeled with the help of simplified chemical models. In these models chemical substances are lumped together forming artificial virtual compounds with virtual characteristics. The paper aims at studying the practices developed in chemical model building and the creation of confidence in these models. Core of the paper will be the analysis of the Urban Airshed Model (UAM) for the Los Angeles region, a pioneering development in the early 1970s. The construction of the UAM involved the "lumping" of chemical processes and extensive testing and tuning. These practices led to a consistent model representation, in which diverse pieces of information fitted and were mutually stabilized. The pragmatic achievement of consistency created confidence, even though empirical tests of the models remained ambiguous and problematic.

When citing this paper, please use the full journal title *Studies in History and Philosophy of Modern Physics*

1. Introduction

Computer models and simulation in the atmospheric sciences involve peculiar features, which have been described as "epistemic opacity", "semi-autonomy" and "simplification". These features have been considered as problematic, because they cause problems of model validation and assessment and involve new forms and not quantifiable ranges of uncertainty of simulation results (e.g. Petersen, 2006). The case of atmospheric chemistry modeling shows that epistemic opacity, semi-autonomy and simplification also provided gains. These characteristics—in spite of being problematic—opened windows of opportunity and created spaces for the realisation of computer models and simulations, which otherwise would have been much more restricted. The construction of the so-called Urban Airshed Model (UAM) to simulate photochemical air pollution in the Los Angeles area provides an interesting example. This model was developed in the early 1970s by a group of researchers around the atmospheric scientist John Seinfeld at the California Institute of Technology.[1] It represented a scientific response to a new type of complex air pollution problem, which was first discovered in the Los Angeles area in the early 1940s.

In this paper I will analyse the construction of the UAM and explore two main issues: first, the practices that have been developed and employed in the construction and use of the Urban Airshed Model, and second, the creation of confidence and trust which the scientists eventually put into the model. The introductory section consists, first, of an outline of epistemic opacity, semi-autonomy and simplification as significant features of computer models and, second, of an introduction into the problem of photochemical air pollution. The core section is dedicated to the analysis of scientific practices developed by Seinfeld and his research group. In the final section I will discuss the characteristics of simulation practices in atmospheric chemistry modeling and the requirements and conditions for the creation of confidence in these practices.

[1] Seinfeld received a Ph.D. in chemical engineering from Princeton University and joined the Department of Chemical Engineering at the California Institute of Technology as Professor of Chemical Engineering in 1967. But he specialized on photochemical air pollution and became known as an atmospheric scientist.

E-mail address: matthias.heymann@ivs.au.dk

1355-2198/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsb.2010.07.002

2. Opacity, semi-autonomy and simplification

Computer models and simulation results need to be tested by comparison with measured data like any physical theory. These model tests are usually called ''model validation'' and—in spite of its limitations—considered crucial for creating trust into the model (Oreskes, Shrader-Frechette, & Belitz, 1994, Oreskes, 1997). But the interpretation of a comparison of simulated and measured data is usually not straightforward and often ambiguous. Testable results of physical theories can be established and scrutinized step by step. Results of computer simulations are constructed in a much less transparent manner due to the complexity of calculation procedures which conceals details of the computational processes. Often, individual elements in simulation models cannot be tested independently, because only their joint net effect is accessible. Humphrey has called this peculiar feature of computer simulation ''epistemic opacity''. ''In many computer simulations, the dynamic relationship between the initial and final states of the core simulation is epistemically opaque because most steps in the process are not open to direct inspection and verification'' (Humphrey, 2004, pp. 147–148). Consequently, it usually cannot with any certainty be decided, whether simulation results fit observational data for the right or the wrong reasons. The confirmation of the validity of a computer model consequently is less straightforward and clear. This observation raises the important question, how scientists develop confidence in their models and simulations (Winsberg, 1999).

Philosophers of science have also described the semi-autonomous character of computer models.[2] These models are based on theories, but involve a complex mixture of heterogenous ''elements'' or ''ingredients'', which cannot be deduced from theory. Model builders take the freedom to draw ''from an astonishingly large range of sources: empirical data, mechanical models, calculational techniques (from the exact to the outrageously inexact), metaphor, and intuition'' (Winsberg, 2003, p. 106). Semi-autonomy similarly holds with regard to the relation of models to real-world-systems they are meant to represent. Due to the lack of experimental data, opportunities of model validation and the establishment of this relation are limited. While this limitation confines opportunities of model testing, it increases at the same time the freedom modelers have in constructing models. The semi-autonomous character of models translates into autonomy of the model builder in constructing his model. This peculiar relation of the limitations of model testing and the freedom of model construction has so far only received little attention in the history and philosophy of computational science.[3] This point is all the more important, as no general prescribed standards exist of what a correct and sufficient model validation is.

Another crucial feature of computer models is the need of simplification.[4] Simplification practices feature in many different ways and create a number of different problems. Computer models based on complex non-linear differential equations, which derive from physical theory, can only be solved with the help of numerical approximation schemes. At least three different forms of simplification have to be distinguished: mathematical approximation, simplification of physical theory and data manipulation. As non-linear partial differential equations are not analytically solvable, they can only be treated with approximate numerical solution schemes, which make use of discretization techniques transforming continuous differential equations into step-by-step algebraic expressions. These mathematical techniques cause round-off errors or instabilities, which however are well understood and can be avoided by keeping to well-defined stability criteria (Dahan Dalmedico, 2001). Numerical approximation involves a discretization in time and space. Continuous processes are represented by defined time steps (like one hour) and by calculation on spatial grids (with one value representing a whole grid element).

Less straightforward is the introduction of simplifications in the physical description of the system to be modeled, which is needed due to limited computer capacity and limited detail of knowledge of the system. Complex differential equations have to be simplified, physical processes or conditions be approximated or neglected altogether and subgrid processes (with a spatial extension smaller than the size of a grid element) to be parameterized (Gramelsberger, 2010; Sundberg, 2009; on parameterization see also Gramelsberger, this issue and Guillemot, this issue). A third type of simplification is involved in the generation of input data and the evaluation of output data. Simulations usually require an enormous amount of input data, which are often not available in desired quantity and quality. Atmospheric and climate models, to take an example, require meteorological data, emission data, chemical and other data with sufficient spatial and temporal coverage. Usually, no such set of data is available. A large fraction of input data has to be constructed artificially prior to simulation, a process which draws from many sources, empirical measurements (as long as such data are available), theoretical considerations, plausible models, educated guesses and a variety of extrapolation and interpolation techniques. Likewise, simulation models produce a vast amount of output data within short time-scales. These data have to be stored, processed and condensed in order to become accessible to interpretation (Edwards, 1999, 2010).

Characteristics of simulation models like opacity, semi-autonomy and simplification are usually discussed in terms of problems. Opacity causes the problem of validation of single elements of the model. No reliable conclusion is possible as to which part of the model caused unsatisfactory results and needs improvement. Likewise, it is not possible to establish, whether results were good for the right or for the wrong reasons. Semi-autonomy is tantamount to the problem that model construction cannot fully be guided by established theoretical and empirical knowledge. Neither can the relation of a model to accepted theory or to the real-world systems it is meant to represent easily be established. It is an open question to what extent scientists have accepted the idea that computer models are not ''realistic'', but pragmatic constructions of virtual processes with rather emulate than represent real processes. Compromising realism is forced by the need of simplification. Scientists have to take decisions, where, how and to what extent best to leave standard physical theory and empirical knowledge, a task for which no clear prescriptions or rules exist.

[2] Morrison and Morgan considered models (not only computer models) as autonomous, because ''models are partially independent of both theories and the world'' (Morgan & Morrison, 1999, p. 10). Winsberg preferred the term ''semi-autonomous'', because partial dependence does not entail total independence, as the term autonomous suggests (Winsberg, 2003).

[3] Humphrey, who so far provided the most extensive investigation of the philosophy of computational science, hardly touches the topic of model validation, but sees ''the danger of a return to a priori science'' (2004, p. 133). Winsberg talks more generally about justification, which can be ''based on considerations coming from theory, from empirical generalizations, from data, or from experience in modeling similar phenomena in other contexts'' (Winsberg, 2001, p. S447). Oreskes et al., 1994 put particular emphasis on the problem of validation (Oreskes, 1997; Oreskes et al., 1994)

[4] Approaches of simplification certainly are no specialty of computer models, but common to all scientific efforts on many levels (Star, 1983). Simplification in the construction of computer models still is particular interest as it involves specific own (and potentially novel) forms and combinations of simplification.