Simulation & Gaming
42(2) 212–224

DOI: 10.1177/1046878110375671
http://sg.sagepub.com

Generation of Evidence in Simulation Runs: Interlinking With Models for Predicting Weather and Climate Change

Gabriele Gramelsberger[1]

Abstract

Meteorology has employed automatic computing machines since the early days of electronic computers. From the 1950s on, a large body of models used for "in silico" experiments (numerical simulation) has been built up, together with an international infrastructure of measuring, modeling, and testing. These outstanding developments—unique in science—led not only to an increasing standardization in developing and applying models but also to deepening the interlinking between modeling and generating evidence. The article explores needs and strategies for evaluating scientific results based on mass data output devices.

Keywords

assessment, climate change, climate modeling, climate policy, computer simulation, computing, evaluation, evidence, interlinking, Intergovernmental Panel on Climate Change, IPCC, John von Neumann, measurement data, model intercomparison, numerical simulation, numerical weather prediction, supercomputers, Vilhem Bjerknes, weather forecasting

The computations of meteorological problems were among the very first computer-based simulated problems using large-scale computing machines (see Goldstine & von Neumann, 1946/1963). In 1948, the meteorologist Jules Charney had conceived

[1]Free University Berlin, Berlin, Germany

Corresponding Author:
Gabriele Gramelsberger, Department of Philosophy and Humanities, Institute of Philosophy, Free University Berlin, Habelschwerdter Allee 30, 14195 Berlin, Germany
Email: gab@zedat.fu-berlin.de

a numerical weather prediction model together with a group of meteorologists at Princeton University and "[Carl-Gustav] Rossby, Vladimir Zworykin of RCA, and Weather Bureau Chief Francis Reichelderfer, had succeeded in convincing von Neumann that weather prediction was a good candidate for his computer" (Phillips, 2000, p. 15). Two years later, the first simulation runs were performed on the Electronic Numerical Integrator and Calculator (ENIAC) and later on the Naval Ordnance Research Calculator (NORC). In 1954, John von Neumann euphorically stated,

> We know today, mainly due to the work of J. Charney, that we can predict by calculation the weather over an area like that of the United States for a duration like 24 hours in a manner which, from the hydrodynamicist's point of view, may be quite primitive because one need for this purpose only consider one level in the atmosphere, i.e. the mean motion of the atmosphere. We know that this gives results which are, by and large, as good as what an experienced "subjective" forecaster can achieve, and this is very respectable. This kind of calculation, from start to finish, would take about a half minute with NORC. (von Neumann, 1954/1963, p. 241)

This statement of John von Neumann is remarkable in the face of an earlier statement made by a meteorologist in 1939.

> "Meteorology is a branch of physics and physics makes use of two powerful tools: experiment and mathematics. The first of these tools is denied to the meteorologist and the second does not prove of much use to him in climatological problems." So many interrelated factors affected climate, he explained, that you couldn't write it all down mathematically without making so many implying assumptions that the result would never match reality. It wasn't even possible to calculate from first principles the average temperature of a place, let alone how the temperature might change in future years. And "without numerical values our deductions are only opinions." (Weart, 2006)

Within one decade, weather prediction has become a computational science producing numerical values to increasingly replace the opinions of experienced subjective forecasters and, later, to evaluate climatological prognoses. The way of producing these numerical values was based on a revolution in meteorology fully introduced in 1904 by the physicist Vilhem Bjerknes (see Bjerknes, 1904).

Bjerknes (1904) articulated a new method of describing weather as a pure mechanical and physical problem. Inspired by electromagnetic phenomena, which he had studied together with Heinrich Hertz at the University of Bonn in 1890, he developed a general circulation theorem for atmospheric phenomena. Conceiving the atmosphere as a problem of fluid dynamics, he described weather in a Newtonian way or as he had entitled his lectures in 1903 at the Stockholm Physics Society as "A Rational Method for Weather Prediction." Bjerknes made use of the 19th-century Navier–Stokes equations that describe the motion of fluids by applying Newton's laws of motion to fluid substances. Instructed by this mechanistic view, he defined the state of the atmosphere by seven atmospheric variables: humidity, temperature, pressure, wind velocity (three scalar quantities), and density. Changes in the atmosphere were defined as the sum of forces acting inside the infinitesimal volumes of the fluid. Using seven equations to predict tomorrow's weather was an elegant way of articulating meteorological problems based on first principles, but these equations were too complex to be solved analytically. Therefore, Bjerknes developed a graphic method to calculate future atmospheric developments based on the current state of the atmosphere measured by, at that time, sparsely installed measurement devices (see Friedman, 1989).

Shift to Computational Methods

In the early 1940s, John von Neumann and Herman H. Goldstine claimed:

> Our present analytical methods seem unsuitable for the solution of the important problems arising in connection with non-linear partial differential equations and, in fact, with virtually all types of non-linear problems in pure mathematics. The truth of this statement is particularly striking in the field of fluid dynamics. Only the most elementary problems have been solved analytically in this field. [. . .] In pure mathematics we need only look at the theories of partial differential and integral equations, while in applied mathematics we may refer to acoustics, electro-dynamics, and quantum-mechanics. The advance of analysis is, at the moment, stagnant along the entire front of non-linear problems. (Goldstine & von Neumann, 1946/1963, p. 2)

Bjerknes's (1904) elegant method has laid down the foundation for modern meteorology, but, at the same time, his method has introduced "modern" problems of computational sciences into meteorology. Replacing unknown analytical solutions of complex equation systems by computation—usually called "numerical simulation"—impose a variety of new practices dealing with computational forms of knowledge production. These practices require huge amounts of measurement data and computations to improve resolution. They face new mathematical constraints. They need new methods of evaluation to achieve evidence beyond the traditional ways of ensuring evidence, for example, by observation. Nevertheless, numerical simulations are the only way to deal with complex systems. Yet at the beginning of the 20th-century, computer power—as well as gathering enough measurement data—was extremely limited, and this, in turn, limited science in general.

Large-scale computing machines were the answer to these limitations, and it is not by chance that John von Neumann was a computer pioneer as well as one of the first experts in numerical solutions of partial differential equations (PDEs). He had become a PDE expert during his stay at Los Alamos before he moved to Princeton University to join the ENIAC team. At Los Alamos, von Neumann had learned that

> the solution of hyperbolic equations with more than two independent variables would afford a great advance to fluid dynamics science. Most problems there involve two or three spatial variables and time, i.e. three or four independent variables. In fact the possibility of handling hyperbolic systems in four independent variables would very nearly constitute the final step in mastering the computational problems of hydrodynamics. (Goldstine & von Neumann, 1946/1963, p. 12)

Therefore, the realization of automatic computing machines exposed computing speed as an important aspect. Vannevar Bush's analogue Differential Analyzer, the only analogy machine von Neumann rated as an acceptable all-purpose device, had to operate 750 multiplications for the calculation of one trajectory. It required 10 to 20 minutes for these calculations for which a human computer would need 7 man-hours. ENIAC needed 2.25 seconds for the same calculations, but its performance was slowed down by storage problems.

Brute force was called for if methods of fast numerical computations were needed. Stanislav M. Ulam (1980) noted at Los Alamos in the mid-1940s, when he teamed up with von Neumann and others: "Proceeding by 'brute force' is considered by some to be more lowbrow" (p. 94). Applied mathematics lacked the elegance and accuracy of mathematical analysis and was called ugly or dull by pure mathematicians. Nevertheless, von Neumann expected the speed of computation would open up new fields of mathematical applications, raising hopes of overcoming hindering stagnation in mathematics and science. For numerical simulations computing speed was and still is decisive.

Today's climate models require several quadrillion operations per simulated year. Current supercomputers calculate 3.5 simulated years per day, but climate models usually cover several hundreds and sometimes thousands of years. Increasing computing power implies more model details, better parametrizations, and finer resolutions of time and space, but it also requires high-resolution measurement data to adequately initialize simulation runs and to evaluate simulation results. Generally speaking, the shift to computational methods—enabled by the mathematization of science, by the increasing use of numerical simulations to overcome the limitations of analytical methods, and the advances in computing performance—has introduced a new way of knowledge production into science with its own practices and epistemic challenges. The development of these practices is still on its way and the epistemic challenges have just appeared on the scientific agenda. One of the most important discussions assesses the evidence that simulation results provide (Thorngate, Tavakoli 2009; Grüne-Yanoff, Weirich 2010).

New Challenges for Generating Evidence

Since Jules Charney's first numerical weather prediction model in 1950, computer-based simulation has become an everyday scientific method of knowledge production. Beside theoretical and experimental departments, computational ones have been established in every scientific research field. Numerical simulations of PDEs as well as statistical and stochastic simulations are increasingly used in physics, engineering, chemistry, and other areas. This ongoing transformation of science into computational sciences poses the question of how simulation-based results can be adequately evaluated or, put differently: Is computational science good science?

Hitherto, evidence has been created empirically by measurement, observation, and experimentation for testing explanations and prognoses and mathematically by proof, deduction, and extrapolation. All these scientific evidence strategies are interlinked with each other and have been refined during the past 400 years. Since the 17th century, empirical and mathematical strategies complete each other and the mathematization of science has dramatically increased. When, in 1848, the French astronomer Urban Le Verrier discovered the planet Neptune, "his discovery was an event of 'pure calculation,' 'the grand triumph of celestial mechanics as founded on Newton's law of gravity'" (Grier, 2005, p. 59). He had numerically analyzed variations in the planets' trajectories by hypothesizing the impact of an undetected planet in orbit around the sun. His prognosis was confirmed by the observatory at Berlin with a single night of observing. The observers wrote to Le Verrier, "'Monsieur, the planet of which you indicated the position, really exists'" (Grier, 2005, p. 60). Forecasting events by calculation is the dream of computational science capable of billions of operations per second today, but evaluating these predictions is not as easy as it was in 1848. Le Verrier had forecasted the impact of the trajectory of a single object based on a relatively simple, that is, highly abstract system—celestial mechanics—dynamically describable and applicable on an observable context, obviously behaving as its mathematical counterpart had forecasted. Nevertheless Le Verrier's discovery was an outstanding result for his time and its by-hand calculations. It has confirmed scientists' trust in the structural isomorphism of mathematical and natural systems. This development has changed the practices of measuring, observing, and experimenting over centuries. They have been increasingly mathematisized and automatisized, producing a proliferating body of data. Justifying a hypothesis by an experiment has turned from a binary yes or no answer into a set of measurement data more or less fitting the mathematically forecasted trajectories of the behavior of a system.

Weather and climate systems as complex systems are based on these new requirements for ensuring evidence. The dynamics of a weather or climate system is defined by seven independent variables, and the space of parameters consists of hundreds of mathematical terms. No meteorologist or mathematician could calculate by hand or oversee by bare imagination the interdependencies of the trajectories that build up a computational weather or climate system. The help of computers as well as algorithms are indispensably needed if meteorologists want to set up a computational experiment. The brute force of computing, if lowbrow or not, creates its own epistemological rules for simulation-based research. The mere quantity of mass data, being used as measurement data for simulation input or produced as simulation output, distinguishes current scientific evaluation practices from 18th- or 19th-century ones. Today's evaluation strategies have to deal with mass data and averaged values. Therefore new strategies have to be developed. Computational meteorology—weather prediction, but in particular climate modeling—provides some interesting ideas for mass data evaluation strategies. Not only is computational meteorology among the oldest computational research fields, because of the social-political requirements, it has become the leading computational science establishing international infrastructures and standards of simulation model development, intercomparison, and evaluation.

In Silico Experiments: Simulation Models and Simulation Runs

Before discussing current evaluation strategies, in silico experimental systems will be explored in detail. Studying the simulation practices of climate modeling unveils the complexity of General Circulation Models (GCMs). Since Jules Charney's simple atmosphere model, meteorology has set up entire earth systems with atmosphere, ocean, land and sea ice, vegetation, and other components. The "engine" of an earth system is still the dynamic core of Bjerknes's (1904) seven equations, a 40-year old piece of code handed down from model to model. The growth of a model is based on the increasing integration of new parametrizations within each component and the coupling of new components.

> Parametrizations are typically based in part on simplified physical models of the unresolved processes (e.g., entraining plume models in some convection schemes). The parametrizations also involve numerical parameters that must be specified as input. Some of these parameters can be measured, at least in principle, while others cannot. It is therefore common to adjust parameter values (possibly chosen from some prior distribution) in order to optimise model simulation of particular variables or to improve global heat balance. This process is often known as 'tuning.' It is justifiable to the extent that two conditions are met:
>
> 1. Observationally based constraints on parameter ranges are not exceeded. Note that in some cases this may not provide a tight constraint on parameter values [. . .].
> 2. The number of degrees of freedom in the tuneable parameters is less than the number of degrees of freedom in the observational constraints used in model evaluation.
>
> This is believed to be true for most GCMs (Intergovernmental Panel on Climate Change–Working Group 1 [IPCC-WG1], 2007, p. 596).

Meteorological simulation models are growing organisms, which age over the years, change, expand, and evolve. These organisms are the result of the collaborative work of two generations of climate modelers. The metaphor organism characterizes these large-scale entities much better than using terms such as *machinery* because they are software entities, not hardware ones such as measurement devices. As software entities, they are extremely flexible, variable, and changeable. Each new mathematical term modifies the whole model and has to be tested. Experiment and experimental setting are tightly interwoven with each other, and both are solely realized by figures. This ontology specifies in silico experiments carried out on simulation models. An experiment based on a simulation model is the performance of the model within specified initial and boundary conditions (simulation run). Depending on the underlying research questions, the same model can be used for various experiments. Differing experiments include modifications of initial data, simulation period, resolution, parameter setting, reference data sets based on measurement for initializing the simulation run (initial data), and the results of interest with respect to research questions, for example, the development of global temperature until 2100 (output data) according to CO_2 input data (hypothesis, e.g., CO_2 doubling in 2020).

A climatological in silico experiment has to define the spatial and temporal resolution of both simulation run and simulation period. Simulation runs are usually based on 16 and more vertical layers and a T42 (distance of grid points projected on Earth scale [equator]: 250 km) or T63 (180 km) resolution of each layer. A T42 resolution creates a global grid of more than 10,000 grid points for each layer and needs a temporal resolution of 20 minutes. Simulation runs and test runs alternate during the experiment. Test runs are based on a higher spatial and temporal resolution and prove the stability of results; for example, T106 (110 km) resolution based on more than 50,000 grid points for each layer and 12 minutes time resolution. Stability is given when a higher resolution improves results within an expected range of values. The simulation period depends on the experiment and the available measurement and reanalysis data. Meteorology has collected reference data sets for various time periods and variables, for example, NCEP-NCAR Reanalysis Temperature Change Plots for 1948-2002, released in August 2003. "The NCEP/NCAR Reanalysis Project is an effort to reanalyze historical data using state-of-the-art models" (National Centre for Atmospheric Research [NCAR], 2008). These data sets are used to initialize simulation runs and to evaluate results under standardized conditions.

All the information of an experimental setting is given as a preset of the simulation run (initialize model, data channels, and parameters; initialize space/time dimension, and memory). After the preset has been completed, the iteration for the first time step starts by initializing the flow of computation, file by file. An atmosphere model consists of hundred of files describing the computation of the model's dynamic and parameterization (diabatic tendencies: cloud cover, radiation fields, vertical turbulence exchange, radiation heating, gravity wave drag, cumulus convection, large scale condensation, land surface processes, lake physics, and update lake/sea temperature, etc.). For each time step (iteration), the simulation run receives data from the ocean model, computes advection and adiabatic tendencies, stores these results as input data for computing the physical parameters (diabatic tendencies), and, after this, goes on computing the horizontal diffusion. Finally, the results are delivered to the ocean model and stored for initializing the next iteration. This procedure has to be performed for each grid point and each time step, for example, for 16 layers × 51,000 grid points (T106) = 816,000 grid points each 12 minutes. A simulated day consists of 120 and a simulated year of 43,800 iterations. It is obvious that the output of such an in silico experiment is a data set of billions of figures based on more than 5 million values of the seven variables—humidity, temperature, pressure, wind velocity (three scalar quantities), and density—describing the state of the atmosphere for each time step at each grid point.

Infrastructure of Model Development, Evaluation, and Intercomparison

Resolving processes grid point by grid point and dealing with billions of figures require new methods of evaluation. Computational meteorology has developed various strategies to ensure evidence for simulation results. These strategies are based on an international infrastructure and standardization for measuring, modeling, and testing. The best way to explore this international infrastructure is the reading of the 2007 published Fourth Assessment Report (AR4) of the Working Group 1 on The Physical Science Basis of Climate Change of the Intergovernmental Panel on Climate Change (see IPCC-WG1, 2007). Chapter 8 of the AR4 addresses "Climate Models and Their Evaluation" and distinguishes between system level and component level evaluation:

> A climate model is a very complex system, with many components. The model must of course be tested at the system level, that is, by running the full model and comparing the results with observations. Such tests can reveal problems, but their source is often hidden by the model's complexity. For this reason, it is also important to test the model at the component level, that is, by isolating particular components and testing them independently of the complete model. Component-level evaluation of climate models is common. Numerical methods are tested in standardised tests, organised through activities such as the quasi-biennial Workshops on Partial Differential Equations on the Sphere. Physical parametrizations used in climate models are being tested through numerous case studies (some based on observations and some idealised), organised through programs such as the Atmospheric Radiation Measurement (ARM) program, EUROpean Cloud Systems (EUROCS) and the Global Energy and Water cycle Experiment (GEWEX) Cloud System Study (GCSS). These activities have been ongoing for a decade or more, and a large body of results has been published (e.g., Randall et al., 2003). (IPCC-WG1, 2007, p. 594)

A typical standard test on the system level requires the proper prognosis of today's climate by using paleo data for initializing the simulation run. Models that fail this test have to be considered as "false." They have to be improved and tested again. Testing models against past (paleo data) and present climate (instrumental data) is the most important part of model evaluation. The evaluation of model data requires fine and homogeneously gridded sets of measurement data for long observation periods but such sets do not exist. Paleo data for climate variables consist of single time series gathered from ice and soil analysis. This source of information is limited in its resolution and its uncertainties are greater than uncertainties of instrumental data. Instrumental data also lack information for wide areas, for example, oceans, and have to be interpolated (data assimilation), but interpolation introduces and distributes uncertainties.

Although the instrumental period goes back to the 18th and 19th century, the distribution of measurement devices has slowly increased. Mass data were not available until satellites were in use. Therefore computational meteorology tries to acquire reference data sets of a fine spatial and temporal resolution for crucial parameters such as temperature or pressure by using methods of reanalysis. Reanalysis is based on a combination of measurement and model data. While simulation models are performed on a homogeneous grid requiring a homogeneous set of data—for initializing as well as evaluation simulation results—measurement is always inhomogeneously distributed. Therefore measurement data sets are incomplete and have to be interpolated by advanced algorithms. These interpolations are based on model data. These reanalysis data sets are internationally used for standardizing experiments (e.g., NCAR, 2008).

Since the late 1980s, two new evaluation strategies have been developed: Model intercomparison and ensemble prognosis.

> The Coupled Model Intercomparison Project (CMIP) is the analogue of AMIP for global coupled ocean-atmosphere GCMs. CMIP began in 1995 under the auspices of the Working Group on Coupled Modelling (WGCM). The PCMDI supports CMIP by helping WGCM to determine the scope of the project, by maintaining the project's data base, and by participating in data analysis. CMIP has received model output from the pre-industrial climate simulations ("control runs") and 1% per year increasing-CO2 simulations of about 30 coupled GCMs. More recent phases of the project include more realistic scenarios of climate forcing. (CLIVAR 2008)

Thirty-five models are involved in the CIMP project. A set of experiments and standardized benchmark calculations has to be performed by each of the participating models. The results are stored by the PCMDI Program for Climate Model Diagnosis and Intercomparison.

> This archive, referred to here [AR4] as "The Multi-Model Data set (MMD) at PCMDI," has allowed hundreds of researchers from outside the modelling groups to scrutinise the models from a variety of perspectives. [. . .] Overall, the vigorous, ongoing intercomparison activities have increased communication among modelling groups, allowed rapid identification and correction of modelling errors and encouraged the creation of standardised benchmark calculations, as well as a more complete and systematic record of modelling progress. (IPCC-WG1, 2007, p. 594)

Ensemble prognoses pay credit to the fact that the average result of a number of simulation runs under slightly changed initial conditions lead in average to better results than the output of each single simulation run. During the last years, metrics have been developed to weight the reliability of model data, but the development is at an early stage. One reason is that ensemble runs require enormous resources of computing power, but computing power is still limited, even though today's supercomputers perform billions of operations per second.

> There is currently no consensus on the optimal way to divide computer resources among: finer numerical grids, which allow for better simulations; greater numbers of ensemble members, which allow for better statistical estimates of uncertainty; and inclusion of a more complete set of processes (e.g., carbon feedbacks, atmospheric chemistry interactions). (IPCC-WG1, 2007, p. 592)

Model intercomparison is a unique evaluation method deployed solely in meteorology. It requires a variety of models as well as an infrastructure of model development, reference data sets, internationally coordinated evaluation projects, and, in general, the international synchronization of main research goals and efforts. Meteorology has built up such an infrastructure during the last century because of the sociopolitical requirements of weather prediction and climate change. Meteorology has also built up a variety of models sharing similar concepts such as the dynamic core but differing in aspects such as parametrization and coupling.

> The future climate change results assessed . . . are based on a hierarchy of models, ranging from Atmosphere-Ocean General Circulation Models (AOGCMs) and Earth System Models of Intermediate Complexity (EMICs) to Simple Climate Models (SCMs). (IPCC-WG1, 2007, p. 749)

No other science discipline exists with such a large body of comparable models and such a coordinated infrastructure for measuring, modeling, and testing. Therefore, computational meteorology is the leading science in domesticating simulation as a reliable scientific method of knowledge production.

Generating Evidence as a Conjoint Endeavor

Testing and evaluating models as well as simulation results has become a conjoint endeavor. For two decades, the international infrastructure of model development, evaluation, and intercomparison has been proliferating: Numerous international working groups, networks, and conjoint projects have been established, interlinking the scientific community of computational meteorology—modelers, model users, and data users. They increasingly build up and publish on community platforms a large body of experience and knowledge, new evaluation methods and practices, and metrics and benchmark calculations, which allow improving reliability of model and model data. This conjoint endeavor is the key in epistemically managing mass data devices such as satellites and simulation models. Its main goal is to analyze the interlinking between modeling and model data but also model data and measurement data by performing simulation runs. Put differently, the main goal is to study model behavior from all perspectives, permanently improving not only simulation models within an international collaboration but also competition. Because of this ongoing endeavor, simulation models have become increasingly reliable for experimentation as well as prognosis. They have become more reliable because the growth of models increasingly involves more realistic aspects of a system. For instance Charney's first weather model was a highly idealized barotropic model where wind develops parallel to isobars. The constraints of his experiments were too strong to perform reliable experimental results, not to speak about prognosis. The first models were conceptual models performed on computers rather than in silico experimental systems as they are today.

The IPCC Assessment Reports give insight into the growth of models and the strategies of dealing with mass data, in particular with in silico results. From this perspective, the reports uniquely document the establishing of an in silico based science, although this process is at an early stage. The IPCC Assessment Reports also show how evidence is generated within the computational meteorology community. Unlike the possibility of failure of real experiments, which falsify the underlying hypothesis because of the ontological difference between theory and experiment, the failure of in silico experiments can have different meanings. It can point to errors of coding, mathematical problems, round-off errors, and other problems, but it does not directly indicate that the hypothesis tested by an in silico experiment is false. The hypothesis of anthropogenic climate change is a good example of this problem. Provided that a well evaluated climate model is used to compute the future global warming based on man-made CO_2 doubling during the next 10 years, the results of increasing mean temperature can indicate correlation with the hypothesis (man-made CO_2 doubling) but does not necessarily indicate the hypothesis' truth. The interrelation between input (hypothesis) and output (results) is a weak one because of the complex interlinking of hundreds of "if, else" conditions of a climate model. No researcher is able to oversee the interdependencies of the trajectories unfolded by the computed model. Evidence is solely generated by the community's conjoint endeavor as an assemblage of knowledge gathered by more than 40 years of practice in climate modeling and simulation. Each simulation run—displayed on community platforms—is deepening the interlinking between the model, its results, and the evidence because each simulation run adds a bit more understanding of complex model behavior and therefore evaluation of the results. Generating evidence is not a simple "yes" or "no" decision or "false" or "right." Generating evidence for in silico based results—in particular for prognoses—is a complex procedure of tests, comparisons, and analysis of the model's behavior—in general (system level) and in detail (component level, parameter level).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

References

Bjerknes, V. (1904). Das Problem der Wettervorhersage vom Standpunkt der Mechanik und der Physik. *Meteorologische Zeitschrift*, 21(1), 1-7. English translation: Bjerknes, V. (2009). The Problem of Weather Prediction, Considered from the Viewpoints of Mechanics and Physics. *Meteorologische Zeitschrift*, *18*(6), 663-667.

CLIVAR. (2008). *The Coupled Model Intercomparison Project (CMIP). CLIVAR Climate Variability and Predictability of the WCRP World Climate Research Project*. Retrieved from http://www.clivar.org/organization/wgcm/cmip.php

Friedman, R. M. (1989). *Appropriating the weather. Vilhelm Bjerknes and the construction of a modern meteorology*. Ithaca, NY: Cornell University Press.

Goldstine, H., & von Neumann, J. (1963): On the principles of large scale computing machines. In A. W. Taub (Ed.), *Collected works of John von Neumann: Vol. V. Design of computers, theory of automata and numerical analysis* (pp. 1-32). Oxford, England: Pergamon Press. (Original work published 1946)

Grier, D. A. (2005). *When computers were human*. Princeton, NJ: Princeton University Press.

Grüne-Yanoff, T., & Weirich, P. (2010). The philosophy and epistemology of simulation: A review. *Simulation & Gaming: An Interdisciplinary Journal, 41*, 20-50.

Intergovernmental Panel on Climate Change–Working Group 1. (2007). *Climate change. The physical science basis of climate change.* Geneva, Switzerland: Author. Retrieved from http://ipcc-wg1.ucar.edu/wg1/wg1-report.html

National Centre for Atmospheric Research. (2008). *NCEP/NCAR reanalyis*. Boulder, CO: Author. Retrieved from http://dss.ucar.edu/pub/reanalysis/index.html

Phillips, N. (2000): The start of numerical weather prediction in the United States. In A. Spekat (Ed.), *50 years of numerical weather prediction* (pp. 13-28). Berlin, Germany: Meteorologische Gesellschaft.

Randall, D. A., Krueger, S., Bretherton, C., Curry, J., Duynkerke, P., Moncrieff, M., . . . Wielicki, B. (2003). Confronting models with data: The GEWEX cloud systems study. *Bulletin of the American Meteorological Society, 84*, 455-469.

Thorngate, W., & Tavakoli, M. (2009). Simulation, rhetoric, and policy making. *Simulation & Gaming: An Interdisciplinary Journal, 40*, 513-527.

Ulam, S. M. (1980). Von Neumann: The interaction of mathematics and computing. In N. Metropolis, J. Howlett, & G.-C. Rota (Eds.), *A history of computing in the twentieth century* (pp. 93-99). New York, NY: Academic Press.

von Neumann, J. (1963). The NORC and problems in high speed computing. In A. W. Taub (Ed.), *Collected works: Vol. V. Design of computers, theory of automata and numerical analysis* (pp. 238-247). Oxford, England: Pergamon Press. (Original work published 1954)
Weart, S. (2006). *The discovery of global warming: Simple models of climate*. Retrieved from http://www.aip.org/history/climate

Bio

Gabriele Gramelsberger is interested in the influence of computation on science and society. During the past 7 years, she has conducted an extensive study on the influence of computer-based simulation on science, in particular in climate research and cell biology. Contact: gab@zedat.fu-berlin.de.