1 Environmental Science Models and Artificial Intelligence

Sue Ellen Haupt, Valliappa Lakshmanan, Caren Marzban, Antonello Pasini, and John K. Williams

1.1 On the Nature of Environmental Science

Environmental science is one of the oldest scientific endeavors. Since the dawn of humanity, along with the ability to reason came the ability to observe and interpret the physical world. It is only natural that people would observe patterns then build basic mental models to predict a future state. For instance, records indicate that there has long been a version of the adage "Red at night, sailor's delight. Red in the morning, sailors take warning."[1] This saying is a simple predictive model based on generations of experience, and it often works. Over time people noted relationships between observations of the sky and subsequent conditions, formed this mental model, and used it to predict future behavior of the weather (Fig. 1.1).

The age of enlightenment following the Renaissance brought a more modern approach to science. Careful experimentation and observation led to uncovering the physics underlying natural phenomena. For instance, a modern understanding of "Red at night, sailor's delight" is based on the theory of Mie scattering. Light rays are scattered by large dry dust particles to the west in the setting sun. According to this theory, large particles tend to scatter the longer wavelength red light forward more than they do the other frequencies of visible light. The long trajectory of the solar beams through the atmosphere when the sun is at a very low zenith angle (such as at sunset or sunrise) compounds this effect. Thus, when light rays from the setting sun are scattered by large dry dust particles associated with a high pressure system to the west, more red light reaches the observer, and the sky appears red. Since prevailing winds in the mid latitudes (where this adage is common) blow from west to east, more Mie scattering at dusk implies that a dry weather pattern is approaching. "Red in the morning, sailors take warning," refers to a similar process at

Sue Ellen Haupt (✉)
Applied Research Laboratory and Meteorology Department, The Pennsylvania State University, P.O. Box 30, State College, PA 16802, USA
Phone: 814/863-7135; fax: 814/865-3287; email: haupts2@asme.org

Valliappa Lakshmanan
Cooperative Institute of Mesoscale Meteorological Studies (CIMMS), University of Oklahoma and National Severe Storms Laboratory (NSSL), 120 David L. Boren Blvd., Norman, OK 73072, USA
Phone: 405-325-6569; email: lakshman@ou.edu

Caren Marzban
Applied Physics Laboratory and the Department of Statistics, University of Washington, Seattle, WA 98195-4323, USA
Phone: 206-221-4361; fax: 206-685-7419; email: marzban@stat.washington.edu

Antonello Pasini
CNR – Institute of Atmospheric Pollution, Via Salaria Km 29.300, I-00016 Monterotondo Stazione (Rome), Italy
Phone: + 39 06 90672274; fax: + 39 06 90672660;
email: pasini@iia.cnr.it

John K. Williams
Research Applications Laboratory, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, USA
Phone: 303-497-2822; fax: 303-497-8401;
email: jkwillia@ucar.edu

[1] For instance, see the quote in Matthew 16:1-2 of the Bible: "He replied, 'When evening comes, you say, 'It will be fair weather, for the sky is red,' and in the morning, 'Today it will be stormy, for the sky is red and overcast' " (NIV).

S. E. Haupt et al. (eds.), *Artificial Intelligence Methods in the Environmental Sciences*

dawn when the low zenith angle in the east would produce more scattering associated with a high pressure system that has already passed, thus suggesting the possibility that a low pressure system is now approaching and wet weather may follow. This example exemplifies the types of physical explanations of observed phenomena that developed in the environmental sciences.

Emergence of modern mathematical techniques gave scientists a new language to describe the natural world. The new physical understanding was codified into partial differential equations (PDEs) that represent the details of the physics. These PDEs can be used to predict the future evolution of a system given its initial state. Modern meteorology began with the development of the primitive equations that describe the conservation of mass, momentum, and energy in the atmosphere. It was necessary to specify the external forcings, including solar insolation and the Earth's rotation. Appropriate boundary and initial conditions had to be applied. Numerical techniques were developed to discretize interlinking equations to form algebraic equations that can be solved simultaneously. In the 1920s, a pioneer of modern meteorology, L.F Richardson, attempted to integrate the full primitive equations by hand. Unfortunately, he did not know about the importance of filtering the equations to avoid the effects of fast gravity and acoustic waves, which caused his integration to "blow up" (Richardson 1922). In spite of the fact that he obtained a very unphysical solution and that the hand integration of the equations took much longer than the weather itself, Richardson made a huge impact on the science of weather forecasting through demonstrating that the equations could be used for prediction and by foreseeing the impact of modern parallel computing. In his 1922 treatise, he imagines an orchestration of human "computers" for numerical weather prediction:

Fig. 1.1(a) "Red at night, sailor's delight." Source: Copyright B.A. Haupt (2005)

Fig. 1.1(b) “Red in the morning, sailor’s take warning.” Source: Copyright B.A. Haupt (2005)

Imagine a large hall like a theatre, except that the circles and galleries go right round through the space usually occupied by the stage. The walls of this chamber are painted to form a map of the globe. The ceiling represents the north polar regions, England is in the gallery, the tropics in the upper circle, Australia on the dress circle and the antarctic in the pit. A myriad of computers are at work upon the weather of the part of the map where each sits, but each computer attends only to one equation or part of an equation. The work of each region is coordinated by an official of higher rank. Numerous little “night signs” display the instantaneous values so that neighbouring computers can read them. Each number is thus displayed in three adjacent zones so as to maintain communication to the North and South on the map. From the floor of the pit a tall pillar rises to half the height of the hall. It carries a large pulpit on its top. In this sits the man in charge of the whole theatre; he is surrounded by several assistants and messengers. One of his duties is to maintain a uniform speed of progress in all parts of the globe. In this respect he is like the conductor of an orchestra in which the instruments are slide-rules and calculating machines. But instead of waving a baton he turns a beam of rosy light upon any region that is running ahead of the rest, and a beam of blue light upon those who are behindhand.

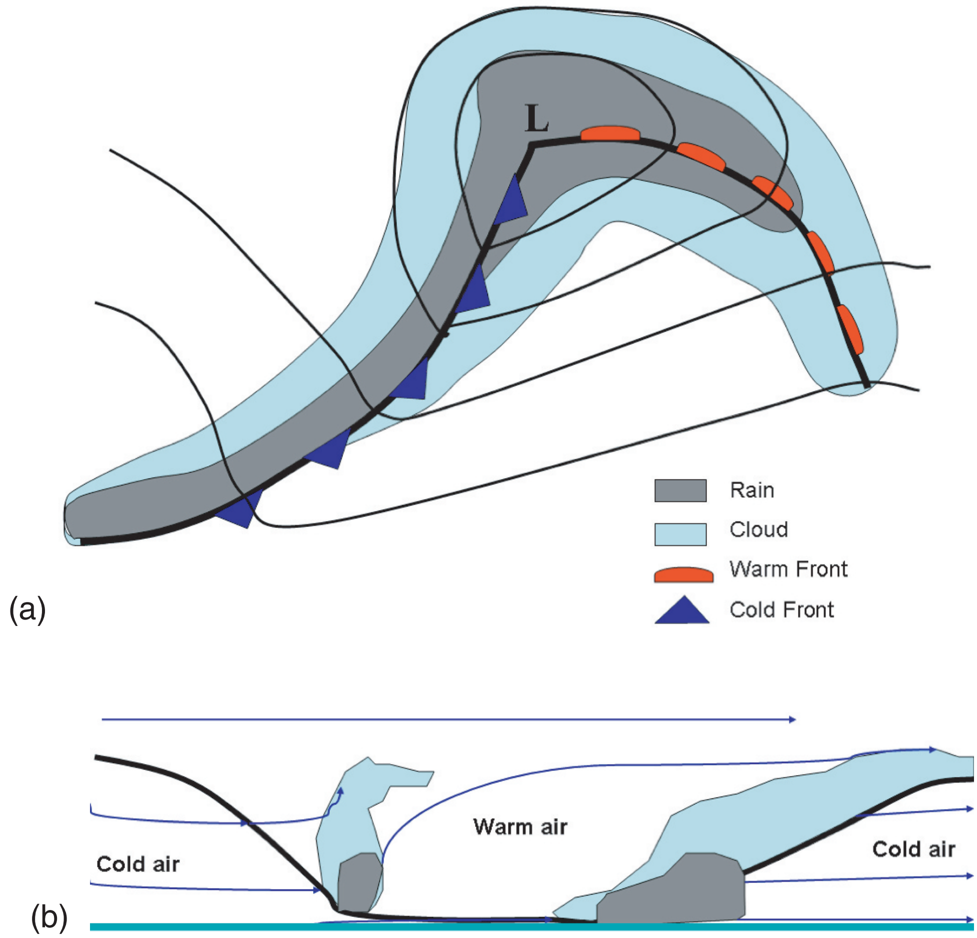

Fig. 1.2 Schematic of the Norwegian Cyclone Model. (a) Horizontal and (b) vertical cross-sections

> Four senior clerks in the central pulpit are collecting the future weather as fast as it is being computed, and despatching it by pneumatic carrier to a quiet room. There it will be coded and telephoned to the radio transmitting station. Messengers carry piles of used computing forms down to a storehouse in the cellar. (Richardson 1922)

At the same time, weather observations continued and more formal models of repeated patterns were formulated. For instance, in meteorology, the Norwegian cyclone model emerged in the early 1920s, attributed primarily to V. Bjerknes and J. Bjerknes (Reed 1977). It describes a weather system in terms of warm fronts, cold fronts, and occluded fronts (see Fig. 1.2). In between the warm and cold fronts is a warm sector. High cloudiness is expected to precede the warm front, which brings in warm southerly winds sometimes preceded by a band of light showers. The warm air rises over both the old cold air being pushed out as well as the new cold air ushered in by the cold front. This conveyer belt of rising warm air produces convection, cooling, condensation, and rain. Convective motion causes showers in the warm sector and deep convection near the cold front often results in violent thunderstorms. So before scientists could accurately model cloud physical processes and their relationship to precipitation mathematically, they could use the Norwegian cyclone model to interpret the skies and predict associated precipitation patterns. Once again, lots of observations came together to form a more formal model of atmospheric phenomena useful for prediction.

With the advent of electronic computers in the 1950s, meteorological research returned to numerical weather prediction, this time aided by the rapid calculations of a machine. The first operational computer, the ENIAC (Electronic Numerical Integrator and Computer) at Aberdeen Proving Grounds became the forecasting tool of Jules Charney, John von Neumann, and R. Fjortoft. They wrote the dynamical relationships in the form of the barotropic vorticity equation, which does not include the fast gravity and acoustic waves that had plagued Richardson. Their efforts met success in 1950 when they produced the first useful numerical weather forecast on the ENIAC (Charney et al. 1950; Lorenz 2006). The field of Numerical Weather Prediction (NWP) was thus born. The field advanced (and continues to advance) through the development of finer resolution models with smaller time steps that can capture more details of the physics. These NWP models have grown to include cloud microphysics, details of radiative transfer, interactions with the biosphere and cryosphere, and dynamic oceanic forcing, among other important dynamical and physical processes. For quite some time, the accuracy of short term prediction continued to improve. Some even thought that if researchers could continue to better define the physics and refine the spatial and time scales, they would eventually be able to perfectly predict the weather arbitrarily far in advance.

This hope was dashed by the famous rediscovery by Ed Lorenz at MIT of what has become known as chaos theory (Lorenz 1963). Dr. Lorenz was studying a version of the atmospheric equations simplified to just three nonlinearly coupled ordinary differential equations (Gleick 1987). He started a numerical integration of his "simple" system of equations then went to lunch. When he came back, his computer run had stopped and he wanted to restart it. He went back several steps and entered the solution at that time as an initial condition for the next run. Then he realized that the numbers coming out of the computer did not quite exactly match those from the succeeding steps of the previous run. Computers were supposed to exactly reproduce their prior computations. What was different? The cause of the disparity was eventually traced to the fact that when he typed in the numbers to restart the program, he did not type in all the digits. The small round-off error grew with time. What he discovered was that in nonlinear dissipative systems, there is sensitivity to initial conditions. This sensitivity means that when a computer model is started with two sets of initial conditions that differ only slightly, the resulting solutions can diverge rapidly in time. Lorenz's observation led to the discovery of chaotic flow and of the limits to predictability in systems of equations such as those representing the atmosphere. What does this imply for forecasting the weather by integrating the dynamical equations? It implies that there is a limit to predictability. No matter how big and fast our computers become, we cannot expect to ever forecast accurately beyond this inherent limit. There is hope, however, that we can find the "strange attractor" of the system, and we are assured that the solution will be somewhere on that manifold of possible solutions. Thus, it might be possible to say something about the likely weather pattern, even without pinpointing the specific conditions. Figure 1.3 pictures the "butterfly"-shaped strange attractor of Lorenz's three equation system in the chaotic regime.

Several new research fields emerged to deal with the discovery of the limits to predictability. The first involves trying to transform atmospheric measurements into an appropriate "initial condition" for a forecast model run. This process is known as initialization. Various complex statistical techniques can be used to first assimilate and interpolate the monitored data to the grid of the model and find a best fit to the steady state model equations. In this manner, the initial condition is tuned to have the lowest possible error to help keep it closer to the solution that could be represented by the model. (See Daley 1991 or Kalnay 2005 for more discussion.)

A second new direction that has emerged is using statistical and data-based approaches to weather prediction. One class of methods is model output statistics (MOS), which compares past model predictions to corresponding records of the observed actual conditions to tune the model or adjust its forecasts (Glahn and Lowry 1972). Another class of methods involves addressing sensitivity to initial conditions head-on by using multiple initial conditions to initialize ensembles of NWP model runs. The forecaster must then interpret the multiple forecasts to make a prediction, but is provided a measure of uncertainty via their variance. At the lowest level, this approach has returned to using the human art of forecasting to distinguish between and synthesize the ensemble of solutions. Of course, it was not long until advanced statistical techniques attempted to emulate that human process by using methods reminiscent of model output statistics to tune the ensemble model runs (Gnieting and Raftery 2005).

Now that statistical techniques have been applied to the forecasting problem, it is natural to ask whether

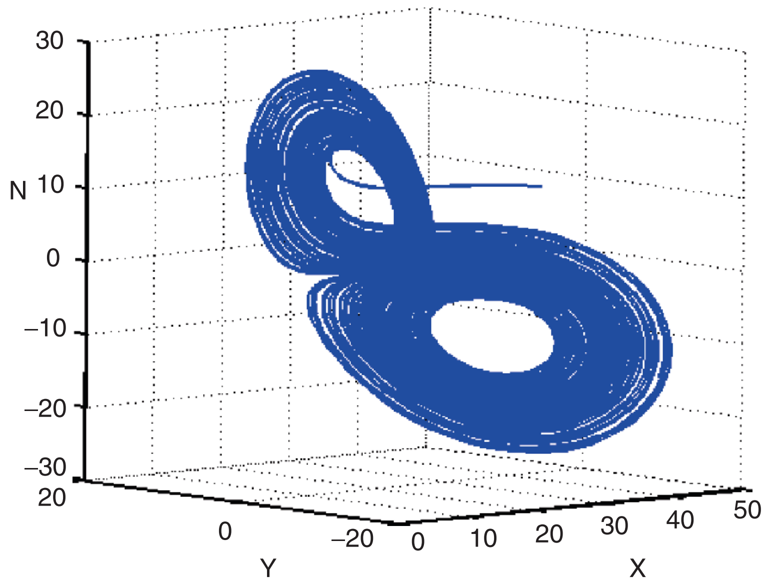

Fig. 1.3 The Lorenz attractor

a completely statistical approach could be developed. Some recent research has attempted to do that to a limited extent. Among the nascent techniques are empirical methods that are no longer based on the deterministic physics, but instead seek ways to glean past information for predictive modeling and to code it into PDEs. Such empirical stochastic methods have been successful at predicting some major atmospheric phenomena such as the El Nino/Southern Oscillation (Penland and Matrosova 1998) and for determining how the atmosphere responds to an imposed forcing (Branstator and Haupt 1998).

These developments leave environmental science in general, and numerical weather prediction in particular, at a crossroads. There are two totally different approaches to forecasting: the first is deterministic or dynamics-based and depends on our ability to write all the dynamical and physical processes mathematically and to discretize them so that they can be solved numerically; the second is empirical or data-based, and depends on the available data and how we choose to use it statistically for prediction. Both approaches typically require large scale computing resources.

Even newer are forecasting techniques that attempt to return to the early methods of recognizing useful patterns without employing formal statistical theory. These methods are empirical data-based methods that do not directly depend on dynamics, but instead seek to model some natural process. The goal of this book is to demonstrate how these "artificial intelligence" methods can be successful at recognizing patterns, performing regression for the purpose of prediction, and optimizing solutions to difficult nonlinear problems. These methods are different from the prior directions of numerically integrating carefully derived dynamical physics equations. Researchers who use artificial intelligence methods are never able to include every tiny detail of the physics in their equations. There is inherent uncertainty in both the measurements and the equations. But looking at how the environment has previously changed will help us predict how it will change the next time, even without modeling all the details. We can use the patterns we recognize, whether they are cold fronts and warm fronts or clustering in some phase space to tell us what typically happens in such a situation. Lorenz's work (1963) implies that we will never be able to exactly predict the future behavior of these complex systems, but it does not inform us whether our dynamics-based methods are likely to be any better than our data-based methods. Many researchers have been highly trained in the physical dynamics-based deterministic methods, and that is certainly helpful background for our adventure into the empirical data-based AI methods.

These developments in weather forecasting are indicative of the types of challenges that arise in the broader field of environmental science. Solid earth geophysicists have explored physical models, data-based models, and artificial intelligence methods to advance their field. So have wildlife biologists, ecologists, water resource managers, air pollution experts, and space scientists, among the many other branches of environmental science.

Many environmental fields deal with complex systems: the atmosphere and the climate system, for instance, or, even more complex living ecosystems and their environment. From a dynamical point of view, we can look at these systems as built on a very high number of component subsystems that interact with each other with feedback. The dynamical approach to grasping this complexity is based on a "decomposition-recomposition" strategy: look at the fundamental subsystems and interactions, phenomena and processes; describe them via dynamical equations; and, finally, simulate the complex system in a computer by means of one or more systems of coupled equations (see Pasini 2005 for an introduction to this dynamics-based approach to modeling in weather and climate studies).

Due to the complexity of the physical systems analyzed, dynamics-based models can suffer from significant problems, such as the difficulty of correctly balancing the "strength" of the many interactions and feedbacks. Often, the modeler must fine-tune the values of the coupling parameters to obtain realistic behavior: obviously, this makes it impossible to uniquely reconstruct the simulated system, thus weakening the reliability of the simulation results coming from the deterministic model. Therefore, the challenge of complexity seems to hint at the need for a more phenomenological approach where the system can be considered as a whole and its behavior can be analyzed in terms of the evolution in time of some representative or important variables of the system, subject to all the interactions and feedbacks. In this framework, artificial intelligence methods have proven useful.

1.2 On the Nature of Artificial Intelligence

Artificial intelligence (AI) has grown out of modern computing combined with a plethora of data to interpret and engineering problems to solve. In some sense, it returns to the earlier methods of analyzing data and trying to build predictive models based on empirical data in a "natural" way. In this sense, AI techniques are typically more data-based than dynamics-based. This use of data can make fast, robust, and skillful forecasts possible in domains that might be intractable by a dynamics-based approach. As Einstein once remarked, "So far as the laws of mathematics refer to reality, they are not certain. And so far as they are certain, they do not refer to reality" (Einstein 1922).

A 19th century scientist would be astonished by the capability of today's computers – solving multi-variable PDEs, running numerical models of the atmosphere, or simulating debris flows, for instance. Yet, that scientist would be equally astonished by the incapabilities of modern computers – their limitations in recognizing a face in a crowd or responding to spoken language, for example. Humans, on the other hand, find that recognizing faces is a cinch but solving multi-variable equations is hard. This is not how it was supposed to be. The earliest computers were billed as machines that could *think*. Indeed, AI has been defined as enabling machines to perceive, understand, and react to their environments. Although perceptive humanoid robots have been a staple of science fiction for many decades, they are still quite impractical. Rather, successful applications of AI have concentrated on single tasks, such as optimally managing the gates at an airline terminal or successfully classifying cells as benign or cancerous.

AI started out by attempting to build upon Aristotelian ideas of logic. Thus, initial research emphasized induction and semantic queries. The goal was to build a system of logic by which computers could "reason" their way from simple bits of data to complex conclusions. After all, humans do this sort of reasoning effortlessly. However, it slowly became apparent that there is more to human reasoning than just induction. Humans, it turns out, are naturally good at many things in ways that current computer designs may never match.

AI researchers scaled back their overly ambitious initial goal to that of building systems that would complement human users and do tasks that humans found onerous. Computers are good at unflaggingly processing reams of data and performing complex computations, but poor at obtaining a holistic view of systems. Humans, good at higher-level thinking, find it hard to do mind-numbing calculations. AI researchers also approached the problem with a new awareness of the impressive performance of biological systems. Thus, many of the new AI approaches were intentionally modeled on the way human experts thought or behaved or on how underlying biological systems such as the human brain worked.

Rather than building an AI system that would replace a human's multifaceted capabilities, researchers concentrated on building special purpose systems that could do one thing well. Thus, instead of building a system that would conduct a wellness check on a patient, for example, they developed a system that would determine an appropriate drug dosage for a cancer patient. The solutions to such targeted problems were called expert systems, because they encoded the rules that an expert in the field would follow to come to his or her conclusions. Computers proved capable of quickly and objectively determining answers to problems where the methodology was precisely defined. The rules in such expert systems are often in the form of decision trees, where the answer to one question narrows down the possibilities and determines what question is asked next, until all possible conclusions but one (or a few) are eliminated.

One fundamental problem in expert systems is how to represent and apply the domain knowledge of experts. Experts often state their knowledge verbally using imprecise words like "not very hot" and "less water." Such words do not lend themselves well to decision trees since the ambiguity can span multiple branches of the tree at each step. This issue is addressed by another AI technique: fuzzy logic. Fuzzy logic provides a framework for encoding imprecise verbal rules – such as those provided by subject domain experts – and aggregating them to yield a final answer. It also allows partial, ambiguous or uncertain evidence to be maintained and efficiently synthesized in the production of the final result. The most celebrated successes of fuzzy logic have been in Japan, where numerous engineering control systems have been built based on encoding expert knowledge in fuzzy rules that are then combined for prediction or control.

Fuzzy logic can be used to create automated decision support systems that model what a human expert would do under similar circumstances. Because humans are considerably skilled at recognizing patterns and often understand the underlying processes that lead to the data, the verbal rules formulated by human experts are often quite robust, even to unseen data and unanticipated situations. A fuzzy logic system, by piggy-backing on such effective analysis, can possess considerable skill. Because fuzzy logic systems are relatively simple to encode and do not require training datasets, they are also fast to create and implement. Another advantage of these systems, often a deciding factor in many applications, is that the fuzzy rules and their synthesis can be naturally interpreted by a human expert. If a training dataset is available, it can be used to optimize the fuzzy logic algorithm's parameters, though this step is not required. Thus, a fuzzy logic system can provide considerable skill and a human-understandable, tunable system for very little investment.

Fuzzy logic systems often provide good solutions to problems for which reliable expert knowledge is readily available and can be represented with verbal rules or heuristics. While there are many situations where this is the case, there are also many others in which either no experts are available or their knowledge cannot easily be represented with verbal rules. One might gauge the suitability of a fuzzy logic approach by assessing whether different experts tend to agree on data cases; if they don't, it might be necessary to code and evaluate multiple fuzzy logic algorithms to represent the range of solution methodologies, which might not be practical. It may also be difficult to determine whether the verbal rules that experts identify is really all that they use to come to their conclusion. Humans often underestimate the role of intuition or the subconscious knowledge brought to bear on a problem. Moreover, many domains in the environmental sciences are exceedingly complex and poorly understood to begin with, so a method capable of automatically recognizing patterns from data may be more appropriate.

Fortunately, another AI method excels at modeling complex, nonlinear systems based on data – the neural network (NN), Like many AI methods, NNs are biologically inspired. The name comes from the fact that they were initially modeled on the way neurons fire, with the accumulated firings of many neurons together determining the brain's response to any particular set of stimuli. The most common architecture used in NNs comprises three layers of neurons – an input layer, a layer of "hidden nodes" and a final output layer. Such an NN can represent any continuous function on a compact domain arbitrarily closely, even a nonlinear one, if it has enough hidden nodes – though choosing the optimal number of hidden nodes for a particular problem may require some effort (Cybenko 1989).

Feed-forward NNs are members of the class of supervised learning machines. In a process called "training", such a learning machine is presented with patterns – sets of inputs and target values, or ideal output corresponding to desired responses to those inputs. The target values may either be provided by an expert in the field, or can be obtained from field surveys, measurements and other information; as such, the target values are often referred to as "ground truth." At the end of training, if all goes well, the learning machine will have created a function that approximately maps the training inputs to the associated targets. Subsequently, when this function is presented with a new set of inputs, it determines a response based on the evidence generalized from the training specimens. Thus, if viewed as an expert system, NNs learn previously unknown relationships or knowledge that experts may not be able to represent with verbal rules. This is because supervised learning machines approximate expert learning behavior, not by approximating the logic that experts use and inexactly describe, but by creating a new mapping to the ground truth associated with the inputs. Specifically, NNs are trained by adjusting their parameters to minimize a cost (or objective) function – a quantity that is usually some function of the difference between the target values and the approximation thereof produced by the network.

Although NNs can represent any continuous function and avoid the problem of depending on expert descriptions by learning data relationships instead, that flexibility comes with a price. First, although the NN learns to approximate the mapping from training samples to target values, the actual function used to represent this approximation is encoded in a large set of connection weights that usually yield no insights. Thus, unlike an expert system, an NN representation is generally not human-understandable, though researchers have found ways to extract approximate rules from an NN in some specific cases (see, for instance, Setiono et al. 2002). The inability to explain in simple terms the behavior of an NN has led to it being called a "black box." However, it is important to point out that this opaqueness is not specific to NNs but applies to many nonlinear models, which may represent the physical world very well but resist being neatly summarized. The fact is that the human desire to explain relationships in simple terms may be inconsistent with the competing requirement to have the most accurate predictions possible, a trade-off that is not peculiar to AI; more details on this will be provided in Chapter 2.

NNs' ability to fit any data places severe requirements on the data necessary for training them. Many NNs have a large number of parameters (weights) that must be estimated during the training phase. The estimates can be highly unreliable if the size of the training data set is not sufficiently large. An abundance of parameters can also lead to overfitting (see Chapter 2), which in turn can adversely affect NNs' performance on "new" data (i.e., not included in the training set). In short, properly training an NN requires lots of data. How much? That question is difficult to answer, but Chapter 2 describes some methods that can help in addressing it.

Another set of biologically-inspired methods are Genetic Algorithms (GAs). They derive their inspiration from combining the concept of genetic recombination with the theory of evolution and survival of the fittest members of a population. Starting from a random set of candidate parameters, the learning process devises better and better approximations to the optimal parameters. The GA is primarily a search and optimization technique. One can, however, pose nearly any practical problem as one of optimization, including many environmental modeling problems. To configure a problem for GA solution requires that the modeler not only choose the representation methodology, but also the cost function that judges the model's soundness. As mentioned above, training an NN usually involves minimizing some cost function, and that process usually requires differentiating the cost function. By contrast, the learning/training process for a GA does not place any restriction on the differentiability of the cost function, so any measure of performance may be used. The GA is also capable of finding optimal solutions to problems such as those in design. Indeed, genetic algorithms may be used to train either an NN or a fuzzy logic system! Using genetic algorithms to train an NN gives us the ability to use non-differentiable cost functions (see Chapter 18), while using GAs to train a fuzzy logic system allows us to improve on human-devised rules by optimizing their parameters.

Another method for inferring the relationship between inputs and targets is to automatically build a decision tree based on the training data set. This approach is also among the fastest in terms of training speed: decision trees can often be trained on substantial data sets in a fraction of the time required by competing techniques. Decision trees, like fuzzy logic systems, also have the advantage of being human-understandable. Unlike fuzzy logic, however, one doesn't need to know the rules beforehand – the rules are learned from training data. Decision trees fell out of favor because the skill realizable with decision trees often lags what is possible using other supervised learning techniques. Recent advances in machine learning – averaging decision trees trained on subsets of the training sets ("bagging") and continually focusing the training on the training data cases that the decision trees get wrong ("boosting") – have made decision trees viable again, but at the cost that the resulting decision trees are no longer human readable. However, aggregate statistics obtained from decision trees are useful in gaining insights into how the decision trees come to their decisions. This is yet another illustration of the aforementioned trade-off between pure performance and transparency.

One of the problems with all of the above data-based methods is that the data on which they are based are always imperfect, corrupted by measurement noise or other artifacts, as are the "ground truth" answers provided in the training data set. Artificial intelligence and statistical methods are closely related in that they both attempt to extract information from noisy data. AI techniques can be utilized to create a practical representation whereas statistical methods can be used to measure how confident we may be that the extracted representation is correct.

1.3 On the Use of Artificial Intelligence in Environmental Science

It is only natural that AI should find applications in the environmental sciences. Let's return to our historical example of weather forecasting, in particular, precipitation forecasting on timescales of about a day. We described a progression from basic generalizations such as "red at night ..." through modern numerical weather prediction with model output statistics and ensemble runs to help deal with inherent error due to sensitivity to initial conditions. What other methods could be used to predict precipitation as well as to address the many other environmental problems?

In Chapter 17, we will see the application of fuzzy logic to analyzing Doppler radar data and to predicting atmospheric turbulence for aviation users. In Chapter 18, neural networks trained with a genetic algorithm will be demonstrated for building models to predict hail. Chapter 11 describes how the radiation physics model in a climate model can be replaced by a much faster Neural Network model with no degradation in the results. The complexity of the Lorenz strange attractor is modeled with a GA in Chapter 18 and an NN in Chapter 12. Some very specific nonlinear phenomena including the El Nino-Southern Oscillation (ENSO) are modeled by NN-based nonlinear principal components in Chapter 8. The use of NNs for assimilating satellite data is developed in Chapter 9 and a specific application described in Chapter 10. Interpreting climate data using an NN is described in Chapter 12. An NN is also used to model the boundary layer height based on radon data in Chapter 13. Chapter 14 describes how a GA is applied to back-calculate the initial conditions of a toxic release if sensor data are available. Chapter 16 discusses advances in image processing techniques through using AI. Habitat suitability modeling using NNs is described in Chapter 19. These chapters sample the utility of AI techniques in a variety of environmental problems.

This book does not try to cover all of the statistical techniques used for environmental study or prediction (there are already excellent entire books on that topic) but instead concentrates on Artificial Intelligence methods. We describe many of these methods and demonstrate their usefulness on some problems addressed by the authors. But we cannot hope to be exhaustive, for it is a very broad field. Similarly, no attempt is made to cover any particular method in great detail. We instead reference the many good treatises that provide details of the methodologies. In attempting to give an overview of many applications, we are unable to provide depth. What we hope to do is to give the reader an introduction to some AI methods and a sampling of the sorts of things that can be done and invite him or her into the field to help make progress in developing and testing alternative methods for modeling the natural world. Plenty of challenges in environmental science have not yet been addressed, and lots of relatively new AI methods could be applied to meet these challenges. The primary purpose of this book is to describe some of the basic

AI techniques, demonstrate some applications in the environmental sciences, and whet the reader's appetite for trying them on his or her own applications. Let the adventure begin.

References

Baer, F., & Tribbia, J. J. (1977). On complete filtering of gravity modes through nonlinear initialization. *Monthly Weather Review*, *105*, 1536–1539.

Branstator, G., & Haupt, S. E. (1998). An empirical model of barotropic atmospheric dynamics and its response to tropical forcing. *Journal of Climate*, *11*, 2645–2667.

Charney, J. G., Fjortoft, R., & von Neumann, J. (1950). Numerical integration of the barotropic vorticity equation. *Tellus*, *2*, 237–254.

Cybenko, G. V. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, *2*, 303–314.

Daley, R. (1991). *Atmospheric data analysis* (457 pp.). Cambridge: Cambridge University Press.

Einstein, A. (1922). *Sidelights on relativity* (56 pp.). London: Methuen & Co.

Glahn, H. R., & Lowry, D. A. (1972). The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology*, *11*, 1203–1211.

Gleick, J. (1987). *Chaos: Making a new science* (352 pp.). New York: Viking.

Gneiting, T., & Raftery, A. E. (2005). Weather forecasting with ensemble methods. *Science*, *310*(5746), 248–249.

Kalnay, E. (2005). *Atmospheric modeling, data assimilation, and predictability* (341 pp.). Cambridge, UK: Cambridge University Press.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, *20*, 130–141.

Lorenz, E. N. (1986). On the existence of a slow manifold. *Journal of the Atmospheric Sciences*, *43*, 1547–1557.

Lorenz, E. N. (1992). The slow manifold. What is it?. *Journal of the Atmospheric Sciences*, *49*, 2449–2451.

Lorenz, E. N. (2006). Reflections on the conception, birth, and childhood of numerical weather prediction. *Annual Review of Earth and Planetery Sciences*, *34*, 37–45.

Lorenz, E. N., & Krishnamurthy, V. (1987). On the nonexistence of a slow manifold. *Journal of the Atmospheric Sciences*, *44*, 2940–2950.

Machenhauer, B. (1977). On the dynamics of gravity wave oscillations in a shallow water model with application to normal mode initialization. *Beitr. Phys. Atmos.*, *50*, 253–271.

Pasini, A. (2005). *From observations to simulations. A conceptual introduction to weather and climate modeling* (201 pp.). Singapore: World Scientific Publishers.

Penland, C., & Matrosova, L. (1998). Prediction of tropical Atlantic sea surface temperatures using linear inverse modeling. *Journal of Climate*, *11*, 483–496.

Reed, R. J. (1977). Bjerknes memorial lecture: The development and status of modern weather prediction. *Bulletin American Meteorological Society*, *8*, 390–399.

Richardson, L. F. (1922). *Weather prediction by numerical process*. Cambridge: Cambridge University Press.

Setiono, R., Leow, W. K., & Zurada, J. M. (2002). Extraction of rules from artificial neural networks for nonlinear regression. *IEEE Transactions on Neural Networks*, *13*, 564–577.