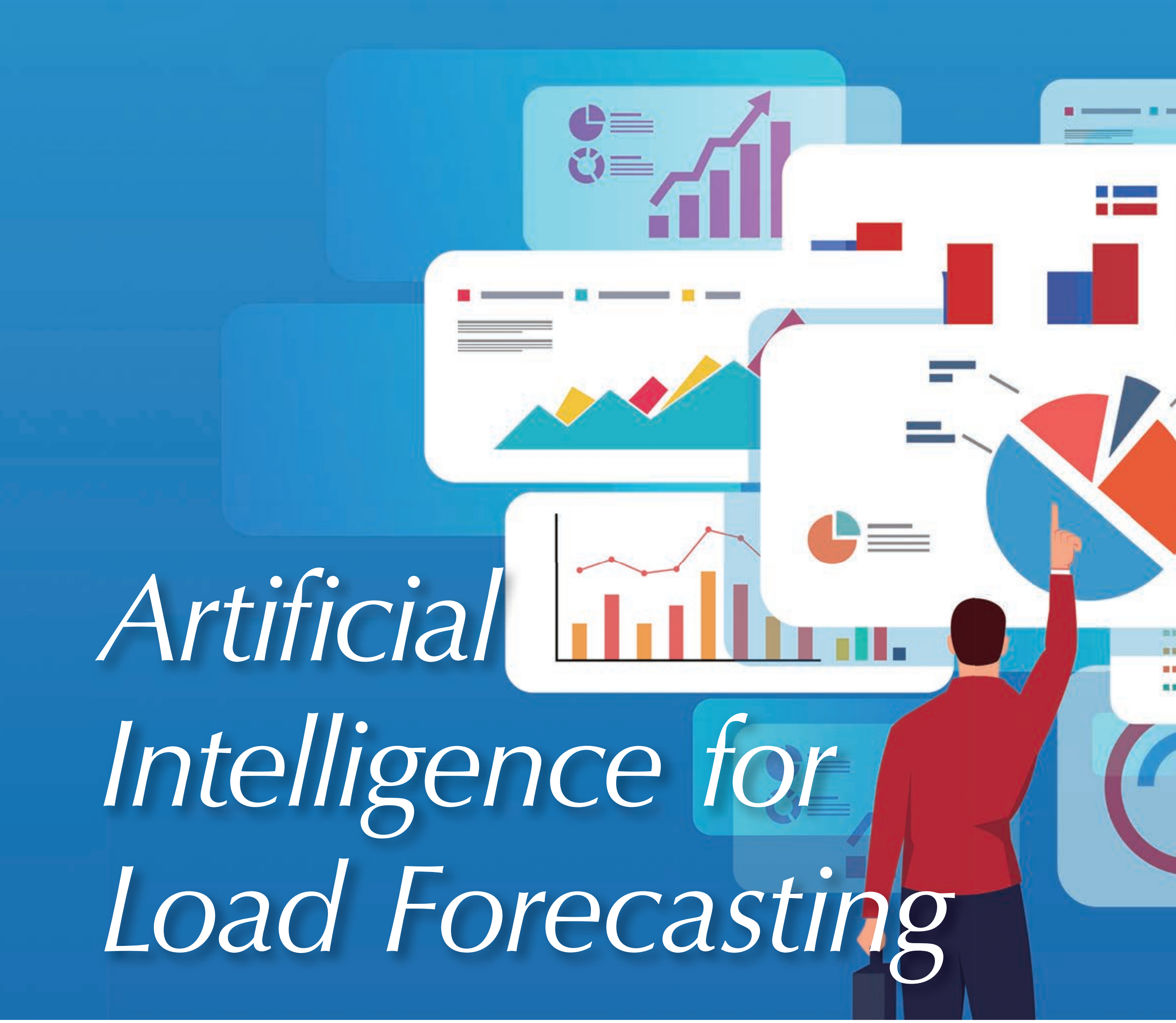

Artificial Intelligence for Load Forecasting

©SHUTTERSTOCK.COM/WAN WEI

History, Illusions, and Opportunities

By Tao Hong and Pu Wang

Digital Object Identifier 10.1109/MPE.2022.3150808
Date of current version: 19 April 2022

ARTIFICIAL INTELLIGENCE (AI) HAS FOUND MANY applications in today's world, such as computer vision for self-driving cars, speech recognition for personal assistants, and algorithm design for strategy gaming systems. Although enjoying the convenience that AI has brought to our daily lives, people may be wondering when and how it started and evolved.

AI has gone through several waves since conceptualization in the 1940s. One of the first commercial AI applications was in the power industry, where artificial neural networks (ANNs) were practically used for short-term load forecasting in the 1990s. During the past three decades, the research community has published thousands of load forecasting papers that promote AI-based models, many of which have originated from or resulted in illusions (misunderstandings,

confusion, exaggerated beliefs, and misleading ideas) of what AI can do.

Today, most of the AI-based load forecasting models proposed in academic literature are still at the theoretical level, with few adopted in practice. In this article, we examine five illusions associated with developing AI-based models. We also present solutions and clarifications to help improve the efficiency of AI for load forecasting, given opportunities in the big data era.

Prepersonal Computer-Era Load Forecasting

After inventing the incandescent light bulb in 1879, Thomas Edison began a series of efforts to commercialize his lightbulb and demonstrate his electric lighting systems. One was the construction of the Pearl Street Station in 1882, the first power plant in the United States, serving an initial load of 400 lamps for 82 customers in lower Manhattan Island. The load profiles during those early years were as simple as step functions, while load forecasting was as easy as counting the light bulbs turned on in the evening. The same method is still being used today by many local distribution companies to estimate the load of streetlights.

The first large-scale, alternating current (AC) electric generating plant in the world, Niagara Falls Power Plant, was built by Nikola Tesla and George Westinghouse in 1895. The AC systems made long-distance power transmission an economically viable solution, which helped power companies realize economies of scale. With larger power generators, electricity could be produced at a lower unit cost. Meanwhile, power companies started to encourage customers to use electricity, further stimulating the invention of electric appliances. As electricity end uses became more diverse, people started to find more curvatures in the load profiles driven by human and business activities, such as turning lights on and off, using electric irons, listening to radio broadcasts, and shopping during holiday seasons. Most of these activities can be captured by calendar variables, just as in today's load forecasting models.

The first modern air conditioner was invented in 1902 by Willis Carrier. Air conditioners not only brought comfort to our daily lives but presented challenges to calendar-based load forecasting methods. As the market penetration of air conditioners increased, the effect of weather started to play a major role in load profiles. In 1944, Henry A. Dryar, a chief load dispatcher at Philadelphia Electric Company, stated "*there is a variable component of the load that reflects the effect of the weather*." He described three weather variables: temperature, wind velocity (wind speed), and cloudiness (cloud cover). Load forecasting then started to attract attention from the power industry and the research community. Figure 1(a) summarizes the important years for load forecasting before the 1950s.

The Electronic Numerical Integrator and Computer (ENIAC), the first programmable, electronic, general-purpose digital computer was made in 1945. Before the wide adoption of personal computers (PCs), many methods were already being applied to load forecasting. Dryar's work represents some of them, assigning weights to different variables to forecast the system load. The concept is like multiple linear regression, which uses several explanatory variables to predict the outcome of a response variable. Apparently, at that time, the parameters were not estimated by computers running statistical software packages.

Some load forecasters kept a book of temperature profiles matched with load profiles for different day types in different seasons. In the morning, they looked at the weather forecast, picked the closest temperature profile from the book, and found the matching load profile for that day. This is also known as the similar day method, which is still used in many

In 1956, a group of scientists participated in the Dartmouth Summer Research Project on Artificial Intelligence, widely recognized as the birth of AI as a research field.

control rooms today, either manually just like in the pre-PC era, or in an automated fashion via computer programs with sophisticated algorithms.

Some load forecasters also figured out other creative ways. For example, a forecaster might take a bicycle ride across the system territory every morning. Upon returning to the office, he weighed his shirt, which carried much sweat during a hot and humid day or little sweat during a chilly and windy day. A heavier shirt indicated a higher peak load for the day and vice versa. This physical method relied on a human body to incorporate the effect of weather on the system load. Today, we can capture weather effects using comprehensive models with various weather variables.

AI's Inception, Two Waves, and Two Winters

Long before the Middle Ages, Chinese, Indian, and Greek philosophers and mathematicians had already begun to investigate formal reasoning. In the 1930s, several mathematicians and logicians attempted to formalize the notion of computability, that is, the ability to solve a problem effectively. In the 1940s, scientists started to investigate whether machines could think like humans. A notable contributor was Alan Turing, who published a series of papers on the topic of intelligent machinery from the late 1940s to the early 1950s.

On the other hand, the foundation of biological neural networks was built in the late 19th century by philosophers and psychologists such as Alexander Bain and William James. Bain and James independently proposed that interactions among neurons within the brain result in thoughts and body activities. The studies of mathematical logic and psychology eventually led to the inception of AI. In 1943, neurophysiologist Warren Sturgis McCulloch and logician Walter Pitts implemented the first artificial neurons that modeled the basic logic units of a brain. A collection of these artificial neurons is called an *Artificial Neural Network (ANN)*, which is inspired by the biological neural networks that constitute animal brains. Figure 2 shows a simple three-layer, feedforward ANN, which has two neurons in the input layer, three in the hidden layer, and one in the output layer. In 1956, a group of scientists participated in

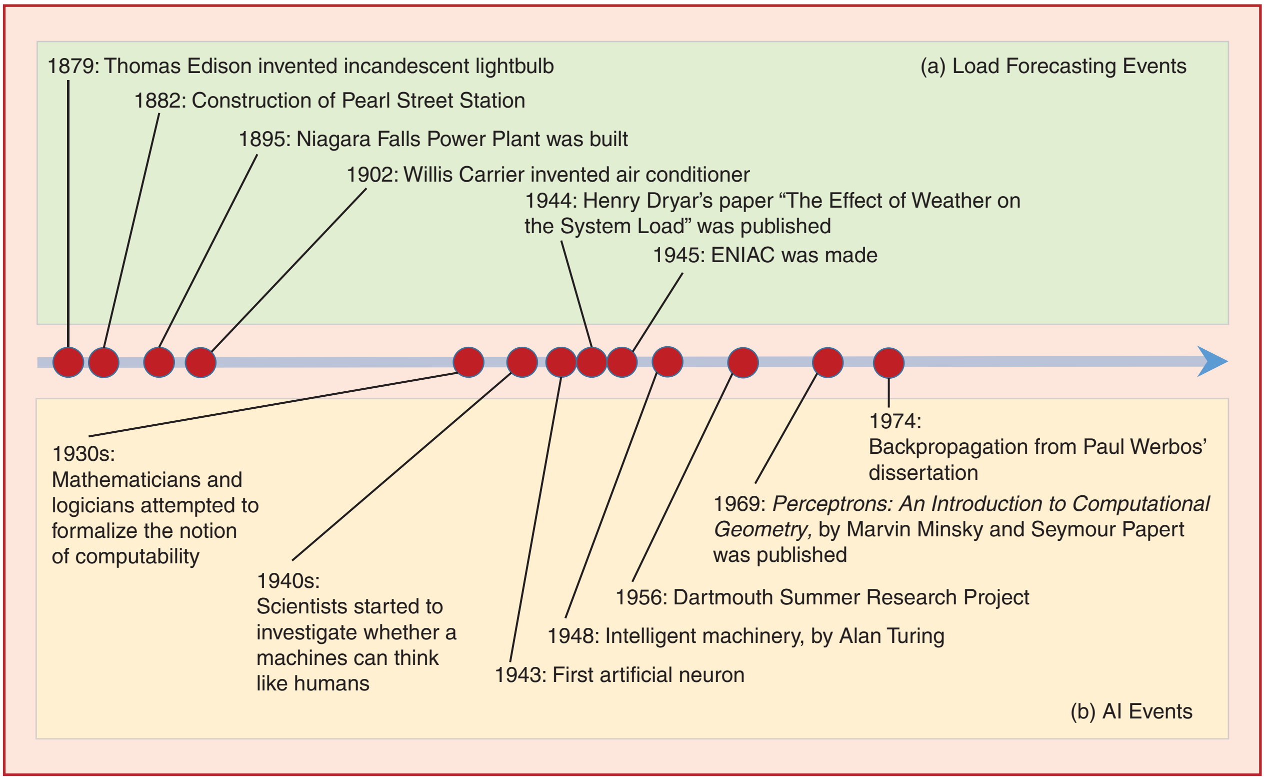

figure 1. The major events related to load forecasting and AI. ENIAC: The Electronic Numerical Integrator and Computer.

the Dartmouth Summer Research Project on Artificial Intelligence, widely recognized as the birth of AI as a research field.

Many theoretical breakthroughs and research advances were made in the 1950s and 1960s. However, the field of AI quickly experienced its first winter after Marvin Minsky and Seymour Papert discovered two drawbacks of ANNs: 1) single-layer NNs were incapable of processing the exclusive-or circuit; and 2) computers were not sophisticated enough to effectively handle the long run time required by large ANNs. The findings were published in *Perceptrons: An Introduction to Computational Geometry*.

The field of AI faced even more challenges in the early 1970s. For example, people realized that many important applications of AI, such as vision and natural language processes, required too much information about the world to be handled by a database at that time. Although some of these challenges were overcome in later years, they slowed down the research progress seriously enough for funding agencies to end financial support.

The second AI wave started in the mid-1970s, when Paul Werbos proposed a backpropagation algorithm to practically train multilayer ANNs. In the 1980s, advances in digital electronics and distributed computing enabled the use of even larger networks. AI saw its first commercial bubble, with hundreds of AI companies being incorporated and millions of dollars flooding into AI research. The bubble, however, burst rather quickly in the late 1980s and early 1990s because commercial vendors failed to develop a wide variety of practical solutions. That was widely considered the second AI winter. Figure 1(b) summarizes the important years for AI before the 21st century.

Applications of AI to Load Forecasting

Despite the second winter AI was facing, power engineers found several AI applications in power systems. In the 1990s, AI became a hot topic among the power engineering community. At power engineering society general meetings, when people asked, "How do we solve 'XYZ' problem?" where *XYZ* might refer to any practical problem such as power quality, security assessment, fault diagnosis, or load forecasting, the answer was often as simple as "Artificial intelligence."

One of the first papers discussing ANNs for load forecasting was presented at the American Power Conference in 1989. Since then, thousands of papers have adopted AI and machine learning (ML) techniques for load forecasting. The topics included fuzzy expert systems, support vector machines, and deep learning. Fuzzy expert systems are based on fuzzy logic, which models logical reasoning with vague or imprecise statements rather than straightforward true-or-false statements. Support vector machines are models that can find a hyperplane in a high-dimensional space to classify observations into one category or the other. Deep learning is based on ANN, where the word *deep* refers to the use of "many" hidden layers in the network.

Figure 3 depicts the number of journal articles through 2021 with "load forecasting" in the title and selected AI/

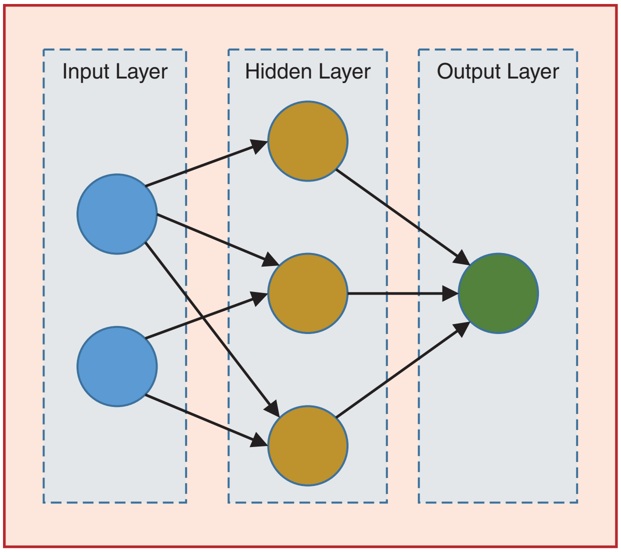

figure 2. A simple three-layer, feedforward ANN.

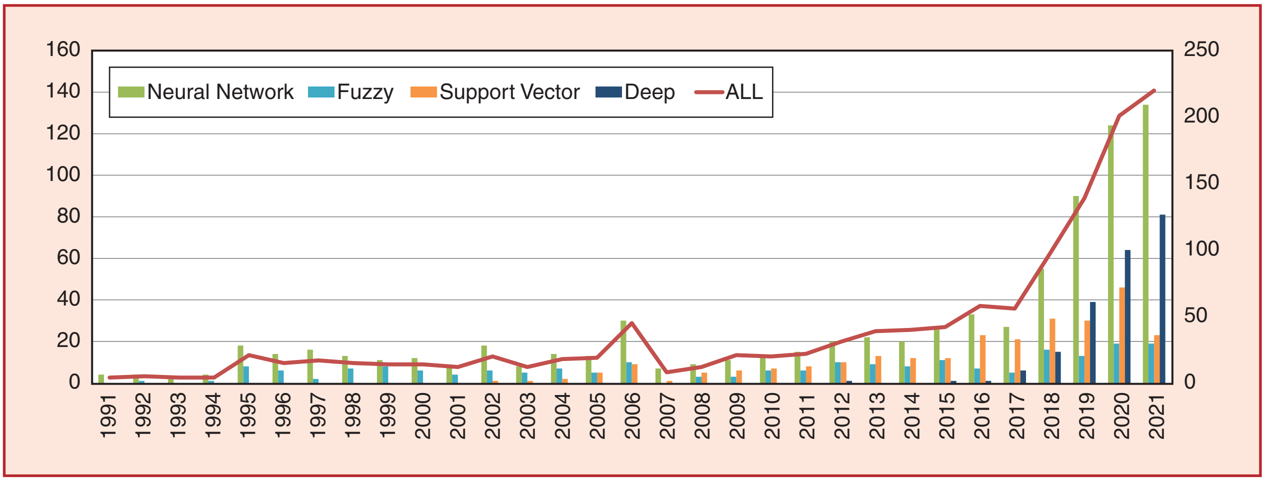

figure 3. The journal papers with "load forecasting" in the title and "AI/ML techniques" in the abstract.

ML techniques in the abstract. In chronological order, load forecasting literature, and journal articles in particular, found ANNs in 1991, fuzzy expert systems in 1992, support vector machines in 2002, and deep learning in 2015. In total, 1,984 journal articles with "load forecasting" in the title were published from 1991 to 2021, of which 1,245 mentioned AI/ML techniques in the abstract. ANNs are mentioned in 796 of them. Indeed, ANNs have been the most popular technique in load forecasting literature during the past three decades.

Most of these ANN models only stayed in academic papers or the laboratory environment. Few received much attention or were accepted by the industry. Nevertheless, one branch of research was able to make load forecasting one of the most successful ANN applications. In 1995, *IEEE Transactions on Power Systems* published an article describing a load forecasting system based on ANNs, which was implemented at 20 United States utilities and used by several of them. The system was later named an *ANN short-term load forecaster* (*ANNSTLF*). In 1997, *IEEE Transactions on Neural Networks* published an article describing ANNSTLF's second generation, with 32 utilities across the United States and Canada using the system. A year later, *IEEE Transactions on Power Systems* published an article describing the third generation of ANNSTLFs used by 35 utilities in the United States and Canada.

Although all three generations of ANNSTLFs are based on ANNs, their structures are quite different. The first generation included 38 ANNs and 24 combiners. Each ANN had a three-layer, feedforward structure, as shown in Figure 2. Some of them had nine input neurons, while others had 72. The 38 ANNs were designed to capture load and weather relationships. The outputs from these ANNs were fed to 24 combiners to generate a forecast for the 24 h of a day. The second generation eliminated the combiners and reduced the number of ANNs to 24, one for each hour of the day. The ANNs were separated into four groups for different periods of the day. The third generation consisted of two ANN load forecasters and one combiner. One ANN was trained to predict the regular (base) load of the next day, while the other ANN predicted the change in hourly load from yesterday to today. The final load forecast for each hour was the linear combination of the outputs from these two ANNs. The commercialized version of the system later became one of the major load forecasting solutions in the power industry, used by many power companies and energy trading firms.

Load forecasting was widely recognized as a successful application of ANNs in the 1990s, but in the following decade, the load forecasting community made little progress in terms of methodological innovation and improvements on model accuracy and practicality. Meanwhile, other communities, such as computer vision and mathematical programming, were making significant progress both methodologically and practically. Looking back at load forecasting literature in the 1990s–2000s and comparing it with those flourishing areas, we can find several reasons for the slow progress:

- ✔ The load forecasting literature had no benchmarking data or models until the 2010s.
- ✔ Many load forecasting papers were not reproducible.
- ✔ Many researchers failed to make direct comparisons with other state-of-the-art load forecasting models.

As a result, load forecasting literature has been filled with papers of varying degrees of quality. Many have created illusions among the load forecasting community, further distracting it from effectively making the next breakthroughs. Some of the aforementioned issues have been or are in the process of being resolved, while others may still take some time to overcome. The remainder of this article will focus on clearing up these illusions.

Modeling Process

Figure 4 presents a typical load forecasting process. At first, a load forecaster would gather load and weather history data to build a load forecasting model. By putting the model and weather forecast together, the load forecaster can generate load forecasts. Some load forecasters may exclude weather data when the load is not weather-sensitive, the weather data are not reliable, or the modeling techniques do not require weather data. Different organizations may

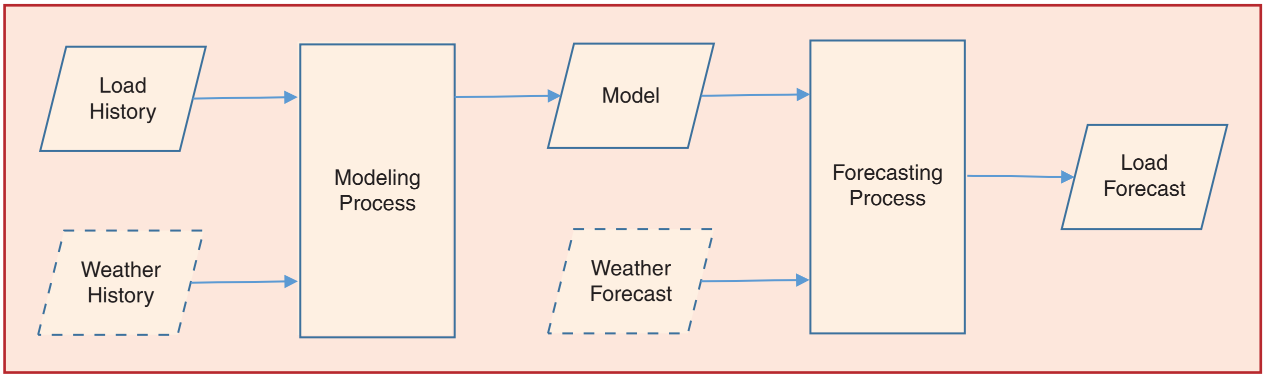

figure 4. A typical load forecasting process.

execute different variants of the process. For example, some may add a preprocessing step to cleanse the load and weather history before sending them to the modeling process, while others may add a postprocessing step at the end to fine-tune the load forecast. Looking at load forecasting literature, most papers have focused on the modeling process by trying various techniques for load forecasting.

For most load forecasting techniques, the modeling processes are somewhat similar. A forecaster slices the historical data into three periods. The first is for parameter estimation (or training). The goodness of fit for this period is called in-*sample fit*. An accurate model would result in a good in-sample fit, but a good fit may not indicate an accurate model. Sometimes a model overfits the data by capturing random noises so that the prediction may be far away from the actual observation or normal energy consumption level. This is known as the *overfitting issue*.

To avoid the overfitting issue, the second period is set for model selection (or validation). The goodness of fit for this period is called a *postsample fit*. A forecaster may train several models, using each to forecast the second period. The model with the best postsample fit result can be selected as the winner for forecasting. In practice, when a large number of models are trained and fed to the validation period, the winning model may be the result of overfitting the validation period.

To add another layer of insurance, the third period is reserved for an out-of-sample test. A forecaster tests the winning model from the postsample fit on the third period, which is completely blind from parameter estimation and model selection. When the selected model does not give surprisingly bad results, it can be promoted for forecasting.

Some variants of the aforementioned process have been used in many business sectors, including load forecasting. For example, the process may be simplified by removing the out-of-sample test. In many case studies, forecasters execute the training and validation steps multiple times on different periods of the historical data, known as *cross validation*. As load forecasting is a time series forecasting problem, sliding simulation is often used in place of training and validation, mentioned previously.

Illusions 1: A Nonlinear Shape, So We Have to Use Nonlinear Models

The relationship between load and temperature is nonlinear, as depicted in Figure 5. The shape of that scatter plot is commonly known as a *hockey stick* or *Nike swoosh*. On the left side, the load increases as the temperature decreases due to heating needs. On the right, the load increases as the temperature increases due to cooling needs.

Many papers used the nonlinear relationship as motivation for nonlinear models. However, a nonlinear relationship does not require nonlinear models. In many situations, linear models are better than nonlinear models at modeling nonlinear curvatures. The "linear" in linear models means *linearity in parameters*. In other words, as long as the parameters being estimated are in a set of linear equations, the model is linear.

The shape in Figure 5 can be modeled by a polynomial regression model, which is in the family of linear regression models. Both second- (blue line in Figure 5) and third-order polynomials (brown line in Figure 5) have been used for load forecasting models. A piecewise linear regression model can also model the shape.

Clarification 1

Linear models can be used to model nonlinear relationships. The business needs designate which family of models should be considered. The exact model to be used should be determined based on the outcome of the model selection process.

Illusion 2: Techniques Are Siloed; Either AI or Statistics, But Not Both

AI and statistics (or econometrics) are often considered rivals. Many load forecasters in either camp rarely look at the work done by the other group. In academic literature, many AI papers compare proposed load forecasting models with other AI-based models. Even if comparisons were made using statistical models, the models are often far less powerful than state of the art in that model family.

ANNs and regression analysis, two seemingly different techniques in AI and statistics, respectively, are connected in many ways. Both techniques ask for similar input variables such as weather and calendar information. Both require the estimation of parameters, which can be formulated as optimization problems. The parameters of ANNs can be estimated through backpropagation, while the parameters (or coefficients) for a regression model can be estimated by minimizing the sum of squared errors. Such regression analyses are called an *ordinary least squares* (*OLS*) *regression*.

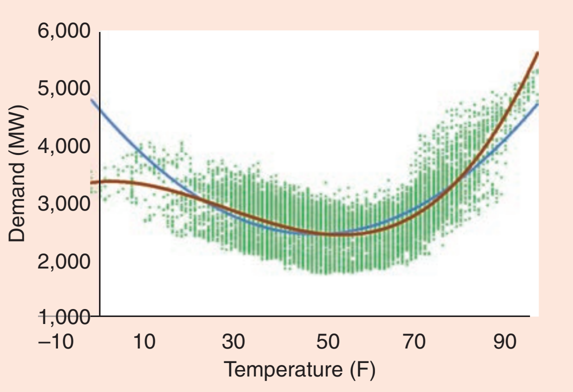

figure 5. Using regression models to capture the load temperature relationship.

Mathematically, the word *norm* can be used to describe the distance of a vector from the origin. Minimizing the sum of squared errors is equivalent to minimizing the Euclidean distance of the error vector from the origin, which is called the *Euclidean norm*, *2-norm*, or L_2 *norm*. Similarly, minimizing the sum of absolute errors is equivalent to minimizing the Manhattan distance of the error vector from the origin, which is called the *Manhattan norm*, *1-norm*, or L_1 *norm*. The regression analysis that minimizes the L_1 norm of the error vector is called L_1 *regression*, or *least absolute deviation regression*.

Starting from least squares regression, we can also obtain several extensions that are well within the field of AI. For example, in addition to minimizing the sum of squared errors, adding a penalty (regularization) to reduce the magnitude of the parameters being estimated would give us what is termed *ridge regression*. Using an extra penalty to reduce the number of variables would give us the least absolute shrinkage and selection operator. Moving absolute errors from the objective function to the constraints section while minimizing the magnitude of the parameters would give us support vector regression (SVR). Imposing a fuzzy membership function to the regression line would give us fuzzy regression.

As listed in Table 1, by simply moving or tuning the norms of both the error vector and parameter vector, we have already connected some of the most frequently used AI and statistics techniques in load forecasting literature. We can treat these decision variables (e.g., which norm to use and where to place them) as hyperparameters and let the model-building process determine the winner. Note that the hyperparameters here refer to the configurations of the modeling process, as opposed to the parameters (coefficients) of a specific mathematical function to be estimated from the data.

There is no clear-cut difference between AI and statistics techniques. Practitioners do not have to separate the two fields either. Connecting them makes the entire flow of research and development (R&D) much more efficient and productive than what has been presented in the literature. For example, when someone discovered that using humidity variables could improve the load forecast accuracy of OLS regression models, a similar benefit would be expected for L_1 regression, SVR, ANN, and so forth. The natural next step would simply be to test humidity variables for the other techniques rather than having another group of people rediscover the same from scratch.

Clarification 2

Many different techniques are connected. Many significant findings, useful insights, and methodological breakthroughs are not technique specific; instead, they are portable among different approaches.

Illusion 3: No One Else Has Used a Combination of These Metaheuristics, So That Is the Novelty

In mathematical optimization, metaheuristic algorithms aim at providing a sufficiently good solution to an optimization problem. They are especially useful when the global optimal solution cannot be easily reached due to limited computational resources. Many well-known optimization algorithms, such as simplex method and interior point method, have specific and strong requirements for the formulation of the problems, so they are constrained for specific types of optimization problems. On the other hand, metaheuristics make few assumptions about the problem formulation, so they can be widely adopted to solve many types of optimization problems.

Many metaheuristic algorithms are metaphor based. For example, genetic algorithms (GAs) are inspired by the process of natural selection, particle swarm optimization (PSO) is motivated by the motion of bird flocks and schooling fish, simulated annealing is caused by the thermodynamic process of metal annealing, ant colony optimization (ACO) is incited by the foraging behavior of natural ants, and gray wolf optimization (GWO) is influenced by the hunting behavior of gray wolves. As various formulations of modeling processes lead to different forecasting techniques, practitioners can conveniently pick some metaheuristics for parameter estimation rather than investigating the specific

table 1. Six variants of regression models.

	Primary Objective to Minimize	Regularization	Constraints
OLS regression	Sum of squared errors	None	None
L_1 regression	Sum of absolute errors	None	None
Ridge regression	Sum of squared errors	Reduce the magnitude of the parameters	None
LASSO regression	Sum of squared errors	Reduce the number of variables	None
SVR	Magnitude of the parameters	None	Upper bound for absolute errors
Fuzzy regression	Spread in fuzzy membership	None	Threshold for fuzzy membership

LASSO: least absolute shrinkage and selection operator.

formulation and its corresponding well-established problem-solving routine, if there is any.

Today, when the metaheuristics community invents a new algorithm, we often see it applied to estimate load forecasting models. In load forecasting literature, we can also find many papers that present hybrid algorithms, such as GA + PSO + ANN, GWO + SVR, and GA + ACO + Fuzzy. The claimed contribution is often twofold: 1) the combination of these specific algorithms/techniques has never been done before for load forecasting and 2) the accuracy from the proposed hybrid algorithm is superior to its counterparts. However, in the authors' experience, most, if not all of these papers tend to be hard to read, impossible to reproduce, and have no presence in the load forecasting tools used by power companies. Although these papers may add load forecasting as an indirect application to corresponding metaheuristic algorithms, their contributions to load forecasting literature appear to be marginal thus far. More about the accuracy issue will be discussed in the "Illusion 5. The Highest Accuracy Reported in the Academic Literature Represents State of the Art" section.

Clarification 3

For most of the models in load forecasting literature, parameter estimation can be formulated as an optimization problem. Metaheuristics is one way to find sufficiently good solutions to these optimization problems. However, combinations or hybrid uses of different metaheuristics with other AI techniques appear to offer a marginal contribution to the load forecasting practice thus far.

Illusion 4: A Three-Layer, Feedforward ANN With 35 Hidden Neurons Is Better Than the One With 41

Researchers have tried many types of ANNs, such as feedforward, recurrent, and deep ANNs. Among the various network structures with different setups, three-layer, feedforward ANNs, due to their relative simplicity, were most frequently used before deep learning was brought to load forecasting. Some papers argue that the superior performance of an ANN structure is due to some minor changes, such as increasing the number of hidden neurons from 35 to 41, increasing hidden layers from one to two, and changing activation functions to radial basis functions. All of these changes can be viewed as hyperparameter tuning. The resulting setup of hyperparameters may work well for one specific data set but fail for the others. The contribution to load forecasting from such results tailored for a specific case study is insignificant.

Nevertheless, this is by no means discouraging practitioners from trying various ANNs or other AI techniques. Instead of devoting resources and efforts to fine-tuning hyperparameters, researchers have been encouraged to look at the bigger picture. Here we list three research directions, which are not meant to be exhaustive:

- ✔ First, investigate novel methods, not a hybrid of different metaheuristics, to efficiently and effectively tune the hyperparameters.
- ✔ Second, find actionable insights, rules of thumb, and practical guidelines to help forecasters better navigate and understand hyperparameter tuning.
- ✔ Finally, explore new structures of ANNs and new AI techniques that may benefit load forecasting. For instance, the structures of the three generations of ANNSTLF are quite different, which led to an improvement in forecast accuracy.

Clarification 4

Given a network structure, such as three-layer, feedforward ANNs, the "optimal" number of hidden neurons is unique on a case-by-case basis. Attention should be paid to exploring the innovative design of ANNs structures, such as recurrent and deep neural networks.

Illusion 5: The Highest Accuracy Reported in Academic Literature Represents State of the Art

Many factors affect load forecast accuracy, such as data quality, load composition, and size of the load. For example, given all other factors set the same, the load forecast accuracy of an independent system operator (ISO) is expected to be higher than that of a power company within the ISO. The accuracy reported in one paper could be for a specific area during a certain period, which may not be generalized to other jurisdictions.

Even putting aside various uncontrollable factors, there may be a tendency for some authors to exaggerate forecast accuracy as a result of flawed processes of model building and forecast evaluation. For example, a paper may report day-ahead forecast accuracy values in 2020 from a particular model M based on a proposed methodology, which beats three other benchmark models based on existing techniques in the literature. All four models are built based on data from 2017 to 2019. Everything may look legitimate, fair, and rigorous on paper. However, behind the scenes, the authors may have tried 10 other variants by tuning hyperparameters of the proposed methodology, among which model M had the highest accuracy. The other variants may not be better than the three benchmark models. In other words, the year 2020 is used for model selection without being disclosed to the readers. This is known as *peeking into the future* in forecasting. In reality, a forecaster cannot pick a model based on perfect information of future load values.

Another approach used to promote a proposed model is cherry-picking. Authors may secretly manipulate data without disclosing details in the paper. Some observations may be excluded from the model-building process, while others may be replaced by different values. All of these changes to the raw data have implications for final forecasts. In other words, data preprocessing methods may have put the proposed approach at an unfair advantage. Meanwhile, readers who do not know these details can hardly make the proposed technique work well in real-world data.

Clarification 5

The accuracy values reported by many academic papers are too good to be true. Some reported values have been significantly exaggerated due to the flawed process of model building and forecast evaluation, such as peeking and cherry-picking. Figure 6 lists a few checkpoints for readers to conduct a smoke test when they see astonishingly good results.

Opportunities

Many opportunities are presented for today's researchers and practitioners interested in load forecasting. Modern computers are more powerful than ever before. In the 1980s, computing resources were a major concern when people developed load forecasting models. Many models had to be built offline and retrained every week or month. The methods at that time had to find a tradeoff between using the most recent data and how fast the models could be estimated. Just like computer technologies pushed AI to its second hype in the 1980s, today's computing infrastructure enables applications of deep learning and other computationally intensive techniques to load forecasting.

- The paper is proposing a mix (or hybrid) of many techniques.
- The paper is merely catching new buzzwords.
- The data is proprietary.
- The paper is not co-authored with industry people (or not sponsored by the industry).
- The proposed method does not utilize new variables
- The proposed method does not take knowledge of other domains.
- The proposed method does not leverage new computing resources.
- The proposed method is dominating its counterparts (another credible method) in all aspects.

figure 6. The smoke tests for too-good-to-be-true results.

In Dryar's 1944 paper, the weather was first identified as a driving factor of electricity demand. In the 20th century, the data used by load forecasters were quite limited. Weather data came from a few weather stations in the service territory. Other than some interval meters collecting hourly demand for selected households for load research, there were no advanced communicating (smart) meters. Utilities had only monthly or bimonthly meter readings for most customers. For long-term load forecasting, monthly or annual load data were used.

Today's load forecasters can access a variety of data sources and high granularity. Smart meters are collecting load readings every hour, if not more frequently. Weather conditions are monitored by thousands of in situ weather stations and satellites. In many places, power companies have installed cameras to capture sky images. Meanwhile, demographics and economic information are monitored and collected in great detail. Social media platforms also generate useful data for load forecasters to better understand human behaviors and local environments.

Forecasting competitions can significantly stimulate the advance of a forecasting field. A well-organized competition can recognize state-of-the-art methods, attract talents from other fields to solve a problem of interest, and release benchmark data for future researchers to continue the investigation.

Table 2 lists five major load competitions during the last three decades. Before the 2010s, load forecasting literature saw only two remarkable competitions. One was supported by Puget Sound Power and Light Company in 1991–1992, and the other was organized by the EUNITE

table 2. Five notable load forecasting competitions from the 1990s to 2010s.

Time	Competition Name	Problem	Affiliations of Winning Teams
1991–1992	Puget Sound Shootout	Short-term load forecasting	University of California, San Diego Southern Methodist University
2001	EUNITE Competition	Medium-term load forecasting	National Taiwan University
2012	GEFCom2012	Hierarchical load forecasting	CountingLab Cambridge University EDF R&D
2014	GEFCom2014	Probabilistic load forecasting	EDF R&D University of North Carolina at Charlotte
2017	GEFCom2017	Hierarchical probabilistic load forecasting	Japan Meteorological Corporation

EDF: Électricité de France.

> The accuracy reported in one paper could be for a specific area during a certain period, which may not be generalized to other jurisdictions.

network in 2001. In the 2010s, a series of three energy forecasting competitions, known as *Global Energy Forecasting Competitions* (*GEFComs*), was organized, attracting hundreds of participants worldwide. These competitions also stimulated interest from the industry, resulting in many other smaller competitions organized by power companies. In the past, good models might be buried by many other mediocre papers. Currently, good models have many opportunities to show up in the leaderboards of various competitions.

Before the 2010s, load forecasting, or energy forecasting in general, was a minority subject in power engineering. Energy forecasters did not have a professional group as their home. In 2011, the IEEE Working Group on Energy Forecasting was formed by the IEEE Power & Energy Society. The group actively promoted energy forecasting by organizing the three GEFComs, delivering panel sessions at major power and energy conferences, teaching tutorials, and so forth. In 2019, the International Institute of Forecasters started the Section on Water, Energy, and Environment (SWEET) to bring together forecasters interested in issues around water, energy, and environment. Today, load forecasters can easily find a welcoming community to network with other forecasters.

Together

Traditionally, the load forecasting job is done by power system operators and planners in power companies. However, the aforementioned load forecasting competitions indicate winning contributions from internationally recognized experts who are not power system operators or planners. Their mastery of techniques from fields such as statistics, econometrics, and computer science allowed them to develop models that proved superior to the competition.

Newcomers to the field of load forecasting often ask, "What is the most accurate technique for load forecasting?" The short and sweet answer should be "None." No technique is superior to all other methods in load forecasting. What matters in load forecast accuracy depends more on the mastery of applying the technique than the technique itself. For example, an ANN expert and OLS regression specialist in a competition may eventually obtain similarly accurate forecasts, while their forecasts are likely to be better than the forecasts developed by people without in-depth experience applying any specific technique or tuning the models.

To accelerate the improvement of load forecasting practices, talented people with experience beyond power companies deserve consideration. As listed in Table 2, the winning teams of those five competitions came from eight different organizations. Only one of the eight organizations is a power company. The other seven include a technology start-up, a weather company, and five universities. Load forecasting is an interdisciplinary subject. Building industry–academia partnerships as well as collaborations among the energy and other communities, such as meteorology and data science, may provide new insights and lead to further breakthroughs. By working together, better models may be realized.

For Further Reading

G. Gross and F. D. Galiana, "Short-term load forecasting," *Proc. IEEE*, vol. 75, no. 12, pp. 1558–1573, Dec. 1987, doi: 10.1109/PROC.1987.13927.

R. Ramanathan, R. Engle, C. W. Granger, F. Vahid-Araghi, and C. Brace, "Short-run forecasts of electricity loads and peaks," *Int. J. Forecasting*, vol. 13, no. 2, pp. 161–174, 1997, doi: 10.1016/S0169-2070(97)00015-0.

A. Khotanzad, R. Afkhami-Rohani, and D. Maratukulam, "ANNSTLF-artificial neural network short-term load forecaster-generation three," *IEEE Trans. Power Syst.*, vol. 13, no. 4, pp. 1413–1422, Nov. 1998, doi: 10.1109/59.736285.

B.-J. Chen, M.-W. Chang, and C.-J. lin, "Load forecasting using support vector Machines: A study on EUNITE competition 2001," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1821–1830, Nov. 2004, doi: 10.1109/TPWRS.2004.835679.

T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," *Int. J. Forecasting*, vol. 32, no. 3, pp. 896–913, Jul./Sep. 2016, doi: 10.1016/j.ijforecast.2016.02.001.

T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int. J. Forecasting*, vol. 32, no. 3, pp. 914–938, Jul./Sep. 2016, doi: 10.1016/j.ijforecast.2015.11.011.

Biographies

Tao Hong is with the University of North Carolina at Charlotte, Charlotte, North Carolina, 28223, USA.

Pu Wang is with Duke Energy Corporation, Charlotte, North Carolina, 28202, USA.