How probabilistic electricity demand forecasts can expedite universal access to clean and reliable electricity

Stephen J. Lee[1,2], Dhruv Suri[3], Priyanshi Somani[4], Christopher L. Dean[2], Jason Pacheco[5], Robert Stoner[1], Ignacio J. Pérez-Arriaga[1], John W. Fisher III[2], Jay Taneja[6]

[1]*MIT Energy Initiative, Massachusetts Institute of Technology*
[2]*Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology*
[3]*Department of Aeronautical and Automobile Engineering, Manipal Institute of Technology*
[4]*Department of Computer Science and Engineering, Manipal Institute of Technology*
[5]*Department of Computer Science, University of Arizona*
[6]*Department of Electrical and Computer Engineering, University of Massachusetts Amherst*

Abstract

The global community is projected to fail in achieving the United Nation's goal of universal access to affordable, reliable, sustainable and modern electricity by 2030. This is ultimately due to inadequate levels of investment. Efforts to right-size infrastructure investments promise improved efficiency: right-sized infrastructure yields more connections with better reliability for every dollar invested. Because of this, geographic information systems, electrification planning models, and methods for characterizing electricity supply and demand have received growing attention as ways to support improved investment decision-making at scale.

In this paper, we highlight an underrepresented and complementary area of research that promises significant value for infrastructure investment decision-making: probabilistic electricity demand forecasting. This paper is organized in three main parts. In the first, we discuss ways in which probabilistic electricity demand forecasts can provide unique value by elucidating economically viable investments that would otherwise be foregone due to misperceptions of their risk. By doing so, probabilistic forecasts have the potential to bolster private investment and expand the resource pool available for electrification and energy for growth. We also discuss how probabilistic forecasts can aid in efforts to efficiently incentivize low-carbon electricity supplies in support of climate goals.

In the second part of this paper, we highlight the fact that probabilistic forecasting models are underrepresented in the literature when it comes to low-access countries. The most prominent forecasts reflect point forecasting methods and demonstrate high variability when compared to historical electricity consumption. There is significant need for probabilistic methods to be applied in this space; probabilistic methods allow for a more informative and transparent way to communicate the expected quality of individual electricity demand forecasts.

In the third and final part of this paper, we highlight a specific model used for probabilistic electricity demand forecasting in the literature: the LDF model. We qualitatively describe useful attributes and limitations of the LDF model and similar probabilistic methods for probabilistic forecasting. We do this with the hope of outlining key modeling concepts that decision-makers should know before employing probabilistic forecasts.

Keywords: energy for growth, electricity access, electricity demand, demand forecasting, probabilistic load forecasting, electrification planning, sustainable development goals, decarbonization, machine learning

---

1. Introduction

There is general consensus that the global community is off-track from realizing the United Nation's Sustainable Development Goal #7 (SDG7) target of "universal access to affordable, reliable and modern energy services" by the year 2030 [1]. Under the International Energy Agency's (IEA) central "Stated Policies Scenario," 660 million people are expected to be without electricity access in 2030 [2]. Rates of improvement are expected to be modest on net, considering that 840 million were estimated without access in 2019 [3]. If we are to

Email address: `leesj@mit.edu` (Stephen J. Lee)

achieve SDG7, the IEA estimates that $35 billion must be spent annually from 2021 to 2030 on generation and network infrastructure towards these ends [2].

Given investment deficits with regards to SDG7, stretching the value of every dollar spent and eliminating inefficiencies is imperative. Right-sizing infrastructure investments to adequately meet electricity demand is central to this endeavor. We can appreciate the value of right-sizing infrastructure by highlighting the risks of over-sizing and under-sizing.

If infrastructure is over-sized relative to demand, customers will be penalized by indirectly paying the costs of generation and network capacity that is never used or will not be used for years. This results in higher costs and disincentivizes consumption. Additionally, because low-access regions experience resource-constraints, overbuilding in some areas means that resources are diverted from underserved areas that would otherwise yield significant benefit from them. In this way, overbuilding makes it so that those without access will remain without electricity for longer periods of time.

If infrastructure is under-sized relative to demand, consumers will experience power outages and industry will suffer. While subsequent upgrades can be made, these investments will miss out on attractive opportunities for exploiting economies of scale in generation and storage, in addition to opportunities for improved network utilization efficiency [4].

Cases of poorly-sized energy infrastructure pervade the sector globally. In 2018, Pakistan paid $4.7 billion in idle capacity charges due to over-sizing [5]. Conversely, from 2003-2006, infrastructure under-sizing led to severe power shortages in China [6, 7]. Even U.S. consumers are affected: generating capacity in the U.S. exceeds required reserve margins by 30%, costing billions per year [8]. In each of these cases, inaccurate electricity demand (i.e. load) forecasts are known to have played a central role [5, 6, 7, 8]. Over-sizing is the direct result of over-forecasting demand, while under-sizing stems from under-forecasting. As such, electricity demand forecasts are key to right-sizing infrastructure.

Empirical evidence for the importance of accurate forecasting is supported by techno-economic models run at multiple spatial and temporal scales [9, 10, 11, 12, 13, 14, 15]. A study employing a building-level electrification model for 366 thousand customers in Uganda demonstrates how electricity unit costs in low-demand scenarios may be nearly three times those in high-demand scenarios resulting from economies of scale and network utilization improvements [16]. Better forecasts allow planners to pursue such economies while minimizing the risk of over-sizing.

While forecast accuracy is critical, forecasting for the power sector and other sociotechnical systems is generally difficult. There are limits to how accurate electricity demand forecasts can be in the face of intrinsically unknown drivers of the power sector, including technological advancements, public policy, a changing population, and climate variability. The future is inherently uncertain.

In this paper, we seek to elucidate the importance of quantifying a forecasting attribute other than accuracy: *forecast uncertainty*. In addition to providing point forecasts that describe what a model perceives to be the most likely future outcome, models can also be used to provide a characterization of its uncertainty (or conversely, its certainty) pertaining to a forecast by way of presenting *probability distributions*. Specialized techniques from statistics and machine learning are well-suited to providing such output as they may learn to characterize uncertainty by evaluating how well historical consumption data fits candidate distributions. Distributions that underestimate uncertainty are penalized as are those that overestimate uncertainty. Fig. 1 illustrates an example of what a probabilistic forecast may look like.

From a power sector decision-maker's perspective, an estimate with high uncertainty (low certainty) could ultimately mean something very different than one with low uncertainty (high certainty), even if the distributions provided center around the same mean forecast. All else equal, decision-makers should make more conservative investments when presented with more uncertain forecasts due to risk-aversion. This may mean investing in smaller generation, storage, and network assets, or investing in grid-compatible mini-grid assets so the option (or "*real option*") is available to connect to the grid at a later date. In other cases, this may mean delaying investments until better information or more resources are available. Conversely, these decision-makers should be more bold when presented with higher-certainty forecasts. They may wish to build larger, sooner, to take advantage of economies of scale that are more likely to bear fruit. Analogous considerations affect infrastructure decision-making around ways to most efficiently meet climate goals under demand uncertainty.

In our view, forecast uncertainty is wrongly under-represented in electricity forecasting and planning communities, especially when considering low-electricity-access (low-access) regions with poor data availability. If forecasts are made, they are too often represented by point-estimates that are likely to engender overconfidence. In this paper, we contextualize the need for in-

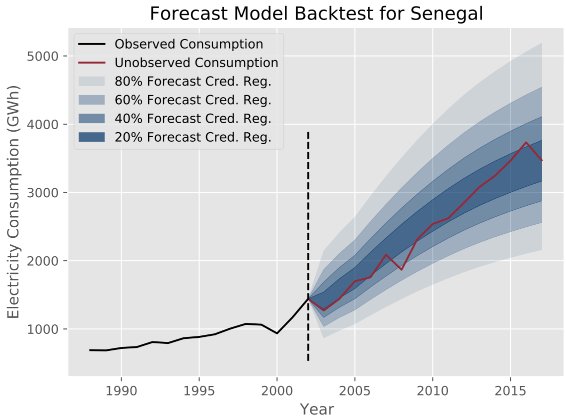


Figure 1: **An example of probabilistic forecasts for Senegal at the country-level.** A backtest for Senegal electricity consumption from 2003 to 2017 is shown. The model observes consumption data corresponding to the time-series in black and provides forecasts with different credible regions as shown in blue. Historical consumption trends that are unobserved by the model are shown in red. Figure from [17] with permission.

creased attention to the development and use of probabilistic electricity demand forecasting (i.e. probabilistic load forecasting) models in low-access settings.

This paper is organized as follows. Section 2 motivates the value that probabilistic demand forecasts can provide in low-access contexts. Section 3 highlights key examples of where probabilistic attributes have gone overlooked in the sector. Section 4 reviews a notable modeling effort that provides probabilistic demand forecasts at the country-level as an example to outline key concepts that decision-makers should know before employing such forecasts.

2. Assessing the value of probabilistic forecasts

Future electricity demand is uncertain because its human, environmental, and technical drivers are themselves uncertain. For instance, we could not have predicted the exact nature of the COVID-19 pandemic and its effects on power consumption years before the first human outbreaks. Neither can we determine when the next major economic recession will occur, what political party will drive U.S. climate policy in five years time, or what battery storage costs will be a decade from now. Because of this inherent uncertainty, we believe that in many cases, electricity demand forecasts should be probabilistic in nature. These characterizations can be informed by data on past and present experiences.

The uncertainty intrinsic to electricity demand forecasting varies by time and space. For example, because aggregate electricity demand generally evolves in a smooth and continuous manner, it should naturally be easier to forecast near-term demand with high certainty than doing so with long-term demand. Another example pertains to regional stability: all else equal, areas with less conflict can be thought to have more predictable demand patterns than those with more.

Better characterizations of the intrinsic uncertainty in future electricity demand can translate into significant bottom-line value if coupled with the right decision-making frameworks. Because probabilistic forecasts and ways to employ them come in various types and qualities, precisely estimating the total economic value that probabilistic forecasting can provide is infeasible. Nevertheless, in this section we outline major ways in which forecasts can provide value, especially when considering electricity infrastructure investments in low-access countries. We expect the potential value from these methods to be significant given the scale of investment needed to achieve SDG7 and subsequent development and climate goals.

2.1. Enabling beneficial investments that could not have occurred otherwise

Perhaps the most concrete way that probabilistic electricity demand forecasts can provide value is by enabling economically viable investments that would otherwise be foregone because of misperceptions of their investment risks. We start this section by introducing key concepts to support the recognition of these benefits. First, we discuss the concept of diminishing marginal utility and explain how it should affect decision-making under uncertainty. We then discuss planning without *any* forecasts and planning with *point* forecasts, before describing how *probabilistic* forecasts can be of value.

2.1.1. Decision-makers under uncertainty should be risk-averse and favor smaller and more modular investments

The concept of *diminishing marginal utility* (DMU) is fundamental to thinking about public decision-making and serving demand for electricity. In this case, DMU reflects the fact that each unit of electricity consumed by a given customer leads to a smaller increase in value relative to that associated with the previous unit [18]. This means that the first units of electricity consumed should bring consumers the most value. In practice, the first few watts of electricity demand may be enough to power basic but critical activities: lighting so that commerce and studies can continue at night, or phone charging so that important communications can be made. Additional electricity consumption is still valuable, but less so on a per-unit basis: consumers with improved supplies may choose to watch television or use electric cookstoves. While economically productive uses of electricity such as irrigating farmland and refrigeration are also valuable, they are highly energy intensive and are only prioritized once adequate electricity supply is available.

Because of DMU, a decision-maker in low-access settings should be more risk-averse when facing demand uncertainty. A risk-averse decision-maker should pursue smaller and more modular investments, all else equal. This is because over-forecasting and over-building entails high opportunity-costs: wasted expenditures could otherwise have gone to electrify or reinforce supplies in other, underserved areas. The net result creates a situation where consumers in some areas are afforded extra capacity that goes unused while consumers in other areas are compelled to live without electricity. Because of DMU, the expected costs from this arrangement outweigh their benefits as high-utility demand is not met. The strategy of making smaller and more modular investments is prudent because, while it may entail the potential for missing out on economies of scale, it decreases the chances of over-building. While chances of under-building are increased, the use of modular supply technologies allow subsequent upgrades and can mitigate potential ramifications pertaining to inadequate supply reliability.

2.1.2. Planning without forecasts

In the absence of any forecast whatsoever, investors are left with only their prior expectations about demand to plan investment decisions. This translates to significant decision-making uncertainty. Given the fact that risk-averse decision-makers with demand uncertainty should favor smaller and more modular investments, only very small investments with economic viability are likely to be made, entailing high per-unit electricity costs. In practice, this may mean an over-reliance on solar kits when underlying demand may instead be able to substantiate much larger minigrid systems or grid connections which derive greater benefit from economies of scale.

Planners can still work to find suitable sites by surveying and collecting better data; however, site-by-site surveys can be both costly and slow, keeping it such that economically viable but unknown investment opportunities remain hidden for undue periods of time.

2.1.3. Planning with point forecasts

It as an improvement for decision-makers to at least be equipped with point forecasts; nevertheless, point forecasts still have notable weaknesses.

Even if point forecasts are accurate *on average*, the uncertainty inherent in sociotechnical systems and limited input information mean that individual point forecasts will necessarily exemplify error. If decision-makers wrongly interpret individual forecasts to be highly certain, they will treat the forecasts with overconfidence when making decisions and will be more likely to overbuild, yielding inefficiencies that prevent other beneficial investments from taking place. As described before, the expected social costs from overbuilding in low-access regions can be pernicious as it means unused supplies are availed in one place at the expense of forgoing higher-value electricity consumption elsewhere. On a society-wide level, these costs are likely to significantly outweigh expected benefits because of DMU.

The situation can be similarly damaging if decision-makers do not know how trustworthy the forecast is likely to be. They may remain too weary to make otherwise sound investments without first still incurring the financial and temporal costs of surveying and data gathering. Economies of scale will be harder to realize if risk-averse decision-makers are too uncertain about forecasts. As in the “planning without forecasts” case, significant value will be left on the table.

Without communicating uncertainty, inaccuracies from outliers in point-forecasts can additionally engender the development of human stigmas that undermine what value point forecasts have the potential to bring.

2.1.4. Planning with probabilistic forecasts

Probabilistic forecasts allow decision-makers to make better decisions in the face of future uncertainty. They can implicitly or explicitly combine these probabilistic characterizations with functions of social utility and attempt to maximize the **expected** utility from their investments [19]. Probabilistic characterizations allow decision-makers to better tune their infrastructure investment strategies, fully accounting for the effects of DMU and economies of scale. All else equal, this should equate to targeting investment projects that provide higher value while also being characterized by lower demand uncertainty. Such projects would be overlooked for less beneficial projects in the absence of probabilistic descriptions. Only when low-risk high-value opportunities are realized should decision-makers choose to make investments with lower value or higher uncertainty.

The value of improved information via probabilistic forecasts also has major implications when considering the importance of 'unlocking' private sector investment in electricity infrastructure for most low-access countries. In these countries, public funding is at all times stretched as far as it can go. In contrast, the private sector represents a comparably boundless source that could manifest reliable electricity connections if viable business cases can be made [20, 21]. Probabilistic forecasts have the promise to illuminate such viable cases and make progress that would not be realized otherwise.

2.2. *Valuing flexibility via real options analysis*

In the previous section, we abstractly introduce how probabilistic demand forecasts can yield improved decision-making considering uncertainty and the effects of DMU and economies of scale. These decisions can be made more concrete by assessing the value of flexibility afforded by some infrastructure classes and not others.

The value of flexibility can be assessed using methods for *real options analysis*. Real options analysis commonly models the present value of the right to make tangible future investments assuming uncertainties in business factors. Part of doing so may entail the definition of strategies to exercise these rights with Monte Carlo simulations drawn from probability distributions over input variables [22].

In the case of electricity demand uncertainty and forecasting, a real options analysis could focus on valuing the flexibility provided by grid-compatible minigrids. In this example, we consider the comparison of two types of investment: grid-*in*compatible minigrids and grid-compatible minigrids. Grid-*in*compatible minigrids involve investment in comparatively cheaper generation, storage, and network components that bring electricity to consumers but do not meet the specifications of the main grid. They entail lower fixed costs relative to grid-compatible minigrids and higher variable costs relative to the main grid. On the other hand, grid-compatible minigrids are more expensive: they ensure that generation, storage, and network components meet grid specifications. All else equal, they entail higher fixed costs relative to grid-*in*compatible minigrids and the same higher variable costs relative to the grid. In a world without the potential for centralized grid connections, it would only make sense to build simple minigrids because of their lower fixed costs; however, grid-compatible minigrids can be advantageous if demand reaches levels that justify connection from the main grid. The main grid can provide reliable electricity with comparatively lower variable costs by taking advantage of much larger economies of scale. If this occurs, assets from a grid-compatible minigrid can be subsumed by the main grid, while those from simple minigrids become obsolete and expenditures in redundant grid-spec assets become necessary. The decision on whether to spend more upfront on a minigrid with the 'real option' to connect to the main grid depends fully on the perceived probability that high demand will eventually warrant connection to the main grid. Better probabilistic characterizations of this demand can support rational decision-making pertaining to such investments in flexibility.

Considerations of flexibility extend beyond the grid-compatible minigrid case. They also apply to possible investments in extensible and modular mini-grid components and solar kits, and more flexible generation and storage assets. Flexibility is also apparent when simply considering the value of doing nothing and waiting for the future when improved information arises about demand, new storage technologies, and other uncertain factors. Understanding the context behind when these technologies should be employed ultimately rests on characterizing probability distributions surrounding business factors including demand. Better forecasts promise to inform strategies for exploiting flexibility and realizing the value of flexible technologies.

In their 2011 book, de Neufville and Scholtes present a simple, contrived example of how uncertainty regarding electricity demand can provide value when building a thermal power plant. Considering variability in demand can help plant managers avoid over-forecasting expected profitability through the probabilistic consideration of potential downside losses associated with non-linear plant operations costs [22]. Agaton and Karl present a more applied example in a 2018 study that uses the real options analysis framework to account for electricity price, oil price, and oil tax externality when making renewable energy investment decisions on Palawan island in the Philippines. Though uncertainties stem from more than just electricity price, and demand uncertainty is not a perfect proxy for electricity price uncertainty, the study is notable because it calculates that the value of the option to invest in renewables can exceed \$150 million on Palawan Island alone [23].

2.3. *Adaptive approaches and the value of information*

With probabilistic electricity demand forecasts, decision-makers can exercise *information planning* along with *infrastructure planning*. Information planning stems from the fact that data provides value, and from the propensity for model-characterized uncertainty to decrease as more and better information is collected and used. Users can assess model-based *value of information* (VoI) metrics and other information theoretic terms associated with different input features of interest [24]. If a piece of information's expected VoI exceeds the practical costs of obtaining it, a modeler would do well to expend resources to procure it. Probabilistic modeling frameworks can rationally direct the calculation of VoI and investments in information alongside physical infrastructure. Moreover, these investments can occur over rolling time frames and continue to adapt as new information is gained, new infrastructure is built, and new revenue is collected. Though we are not aware of empirical examples for which such adaptive approaches to electricity infrastructure planning have been exercised at scale, the concept of *Adaptive Electricity Access Planning* has been proposed in [25].

2.4. Climate goals and low-carbon infrastructure planning

Global climate efforts, goals, and agreements add an additional layer of complexity to electricity infrastructure planning. Researchers are projecting that achieving the Paris climate goals will be unlikely given current rates of progress [26, 27]. The electric power sector is central to these goals. In 2014, electricity and heat accounted for 25% of global greenhouse gas emissions worldwide [28]; however, decarbonizing broader sectors of the economy will very likely necessitate expanding the scope of electrification and meeting new demand with renewable supply [29, 30].

Probabilistic characterizations of future demand in low-access regions can be used to efficiently achieve low-carbon electricity supplies in much the same way they can improve infrastructure right-sizing efforts. Two key concepts are useful to define before supporting this claim: *energy system momentum* and *committed emissions*. Energy system momentum reflects the fact that long-lived generation, distribution, and transmission assets, in addition to the regulatory frameworks that govern them, yield significant systems-level inertia. In the context of climate, energy system momentum yields committed emissions. Already-purchased physical assets reflect sunk fixed costs and are thus advantaged relative to assets that are yet to be bought. Fossil fuel-fired generators can be thought to have future emissions effectively embedded within them [31]. Because of this, decisions made now affect the future carbon intensity of the sector for decades by way of committed emissions.

These effects inform the planning problem for low-carbon energy systems in low-access contexts. Consider a situation in which a low-access country with electricity demand uncertainty is committed to reaching some predefined level of emissions by a future year, such as from a Nationally Determined Contribution (NDC). If future demand is uncertain, the target shares of renewable and carbon-emitting electricity generation will also be uncertain. Too much upfront investment in fossil-derived generation capacity and too little investment in renewable capacity can result in futures for which it is overwhelmingly costly and inefficient to achieve climate commitments. Too little upfront investment in fossil-derived sources and too much investment in renewables can yield electricity supplies that, in the short-term, are overly costly and unreliable due to the intermittency of renewables. Understanding demand uncertainty can help risk-averse decision-makers target balanced levels of renewable and fossil-derived generation investment over time; this allows for improved risk management in meeting emissions targets while seeking to maximize economic efficiency and growth.

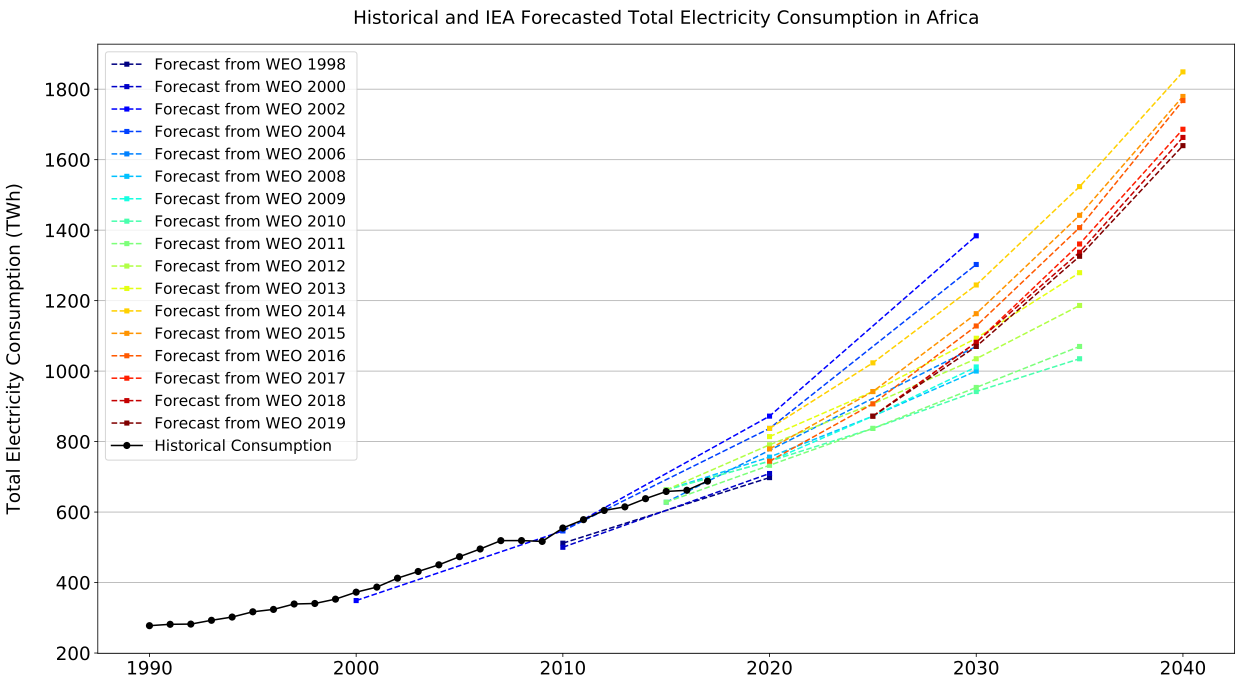
Characterizations of demand uncertainty can also aid in the process of *defining* emissions targets such as NDCs. If future demand is uncertain, then there are higher chances that without uncertainty characterization, risk-averse decision-makers will define overly-restrictive targets that can stymie growth in the power sector and in the economy. If future demand is more certain, appropriate emissions targets can be used to more effectively and equitably reduce carbon emissions worldwide.

3. Probabilistic forecasting is underrepresented in the literature

Despite the potential value conferred by probabilistic electricity demand forecasts, these models are seldom employed in low-access contexts.

A 2020 review paper focusing on electricity demand forecasting in "low and middle income countries" by Mir et al. exemplifies this point: while the review details the importance of accurate forecasts, surveys all of the major model classes used to forecast electricity demand, and cites over 130 articles, the authors provide no description of the difference between a probabilistic forecast and a point forecast, nor of the former's relative merits [5].

In contrast, a 2016 review paper focused exclusively on probabilistic electricity demand forecasting by Hong and Fan review common methods, describe notable studies, and characterize the frontier of research in this subfield (i.e. the need to apply probabilistic forecasting



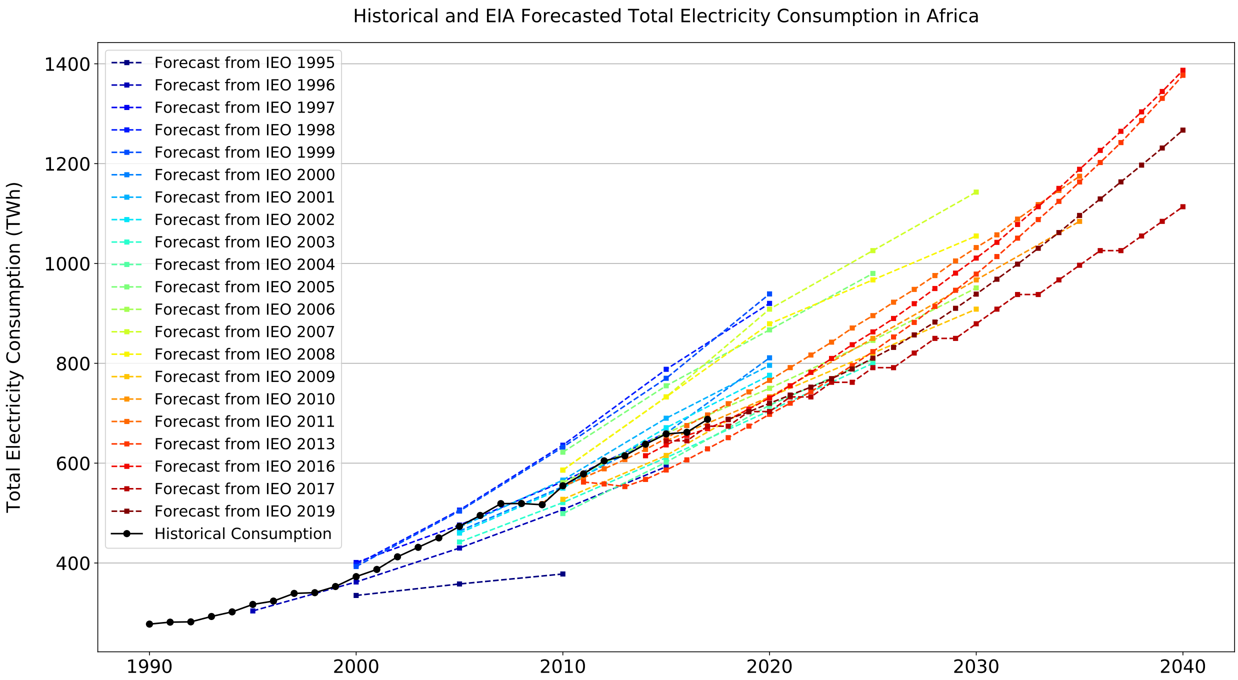


Figure 2: **Historical and forecasted electricity consumption** are shown from the IEA's WEO series (top) and the EIA's IEO series (bottom).

models to account for climate variability, electric vehicles, demand response, energy efficiency, and wind and solar power generation) but fails to mention the specific applicability and need for probabilistic demand forecasting in low-access contexts [32].

3.1. A mini-review

We share a table detailing our own, non-exhaustive literature review in Table 1 in the supplement of this paper. We searched for electricity demand forecasting papers in Africa, specifically. Critically, there are very few studies that provide probabilistic outputs; the vast majority of forecasts published reflect point-estimates. With one notable exception, all of the studies providing probabilistic outputs exclusively focus on the country of South Africa. The fact that South Africa is often considered a high-electricity-consumption outlier on the African continent underscores the modeling gap that exists regarding probabilistic forecasting for low-access countries.

We highlight the one notable exception to this trend in the first row of Table 1. It is actually a paper we wrote as a technical companion to this specific paper [17].

Section 4 shows how [17] uses the lightweight data fusion (LDF) methodology [33], along with long short-term memory (LSTM) models, to provide and validate probabilistic forecasts. We speculate that one of the reasons that probabilistic electricity demand forecasts have been slow to provide answers pertaining to low-access countries and SDG7 is because of general data availability issues. Even aggregated country-level forecasting was made challenging by inconsistent data ranges and data availabilities across key features of interest [17].

3.2. A deep dive into prominent demand scenarios

In this subsection, we review electricity demand scenarios provided by the IEA's World Energy Outlook (WEO) [34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50] and the U.S. Energy Information Administration's (EIA) International Energy Outlook (IEO) [51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71] series and we compare projections by previous editions to actual historical demand for Africa.[1] Data series used from the WEO specifically corresponds to the "Stated Policies Scenario" (or equivalently, the "New Policies Scenario") which reflects both existing and announced policy frameworks and intentions. Data from the IEO corresponds to the EIA's "reference case," reflecting current and anticipated trends. The IEA's scenarios result from the World Energy Model (WEM) [73], while the EIA's forecasts are driven by the World Energy Projection System (WEPS) model [74].

Both the IEA and EIA provide data and documentation associated with their modeling frameworks. The WEM is a simulation model that links supply and demand across different sectors of the economy to outline scenarios of future energy flows, CO_2 emissions, and investments. In determining electricity demand, it uses econometric methods to relate historical data and exogenous assumptions on socioeconomic drivers to determine demand-side drivers (e.g. steel production in industry, household size, etc.). It then uses a least-cost approach to determine fuel and technology type allocations and accounts for efficiency levels to determine final demand levels [73]. The WEPS system is comprised of a set of models that simulate the international energy system, including models for global output, residential demand, commercial demand, industrial demand, and transportation demand. These models use dynamic econometric equations, ordinary least squares and least absolute difference regression, model selection algorithms, and bottom-up approaches to determine demand from different segments [74]. Despite the multi-faceted and complicated nature of these systems, when considering Africa, neither the IEA nor EIA provide historical scenarios with country-level resolution and instead only provide electricity consumption figures for the whole continent.[2]

To get a sense of the success of their scenarios, we choose to investigate the accuracy of the IEA and EIA's electricity scenarios at the continent-level via backtesting. Similar backtesting analyses pertaining to total primary energy consumption was previously published by Wolfram [75]. Here, we compile data from IEA's WEO between 1998 and 2019, and U.S. EIA's International Energy Outlook between 1995 and 2019. Fig. 2 shows deviation between the forecasts and historical electricity demand. Colored dashed line segments depict historical forecasts, while the solid black line depicts actual consumption values on the continent.

From Fig. 2 it is apparent that both the IEA and EIA

[1] While neither the IEA or EIA claim that their scenarios are meant to be directly interpreted as "forecasts," the "outlooks," "futures," or "scenarios" they share are commonly interpreted and used as such [72]

[2] Recent editions of the WEO and associated documents by the IEA have started to share country-level scenarios, but only for select countries. These scenarios are also too recent to enable backtesting.

commonly change their point-estimates from year-to-year and these estimates collectively convey high variability and uncertainty at longer time horizons. Additionally, demand growth rates have systematically been underestimated in recent forecasts. Such historical characteristics underscore the need for forecasts in the form of estimated probability distributions. Figs. 5, 6, and 7 in the supplement reflect analogous analyses performed for India, China, and the United States.

4. Key concepts related to machine learning-based probabilistic forecasts

In this section we give a concrete example of what probabilistic demand forecasts may look like and outline key concepts that decision-makers should know about the model classes employed to produce these forecasts. In particular we highlight a recent study [17] that provides 15-year annual probabilistic electricity demand forecasts for 52 of 54 African countries. This study uses the lightweight data fusion (LDF) methodology, originally developed in Dean et al. [33], that combines probabilistic latent variable models and neural networks to learn from multiple data sources. At a high level, LDF models a probability distribution over electricity demand and learns a mapping from related features—e.g., historical electricity consumption, GDP per-capita, population, cooling degree days, heating degree days, battle-related deaths, etc.—into the parameters of that distribution. The instantiation of LDF in [17] uses a long short-term memory (LSTM) network as the mapping function and a gamma-Poisson latent variable model; for specific modeling details see [17, 33].

Fig. 1 shares a historical forecast made for Senegal. The model observes 15 years of input data between 1988 and 2002 (inclusive). The black line shows historical "observed" consumption data. The model then produces probabilistic forecasts for every year between 2003 and 2017 (inclusive) as shown using blue credible regions. The 80% credible region reflects the smallest interval where the model believes future consumption will be with 80% certainty. 60%, 40%, and 20% credible regions are also plotted. Because the specific experiment shown constrains the model to never observe historical consumption data between 2003 and 2017, we can use such unobserved historical consumption data to validate the performance of the probabilistic forecasts. Unobserved historical consumption is plotted with the red line.

4.1. Useful attributes of probabilistic machine learning forecasts

As discussed in Section 2, probabilistic forecasts can enable beneficial investments that could not have occurred otherwise, enable the valuation of flexibility, and provide a framework for adaptive information and infrastructure planning systems. In this subsection, we take a deep dive into useful features of many probabilistic forecasts that decision-makers should know about to better understand and use such forecasts.

Because we specifically focus on the LDF model [17] in this section, some of the topics discussed will not be applicable to all probabilistic forecasting methodologies. Nevertheless, they should be common to many machine learning models for probabilistic forecasting, especially those that follow a *Bayesian framework* for variations of *supervised learning*. Bayesian methods rely on Bayes' Theorem which provides a way to calculate the probability that an event occurs using prior knowledge and observed data. Models employing supervised machine learning use observed input-output pairs to learn a function capable of mapping previously unseen inputs to outputs; in the application of LDF presented, electricity consumption values in forecast years are used treated as labels to inform output probability distributions for electricity consumption.

4.1.1. Probabilistic machine learning models generally follow a more inductive, as opposed to deductive, epistemology

Fitting probabilistic forecasting methods that follow a supervised approach necessitate historical data. As such, we can say they follow an *inductive* epistemology: the models begin with observations of reality and learn patterns that can be tested in practice. These methods make an implicit set of assumptions: that there is a relationship that can be learned between future electricity consumption and historical data on all of the input features employed, and that this relationship persists through the dates considered. The model interprets these relationships between historical inputs and trends fully automatically, requiring limited human input. Human input is strictly optional: users provide prior beliefs about forecast distributions only if desired.

In contrast, some forecasting methods, including the IEA's WEM model [73] and the EIA's WEPS model [74], are systems that rely very heavily on theory and deduced hypotheses. As such, we can say that they follow a more *deductive* epistemology. In these systems, modelers define specific ways in which different sectors interact, how exogenous forecasts of input variables are

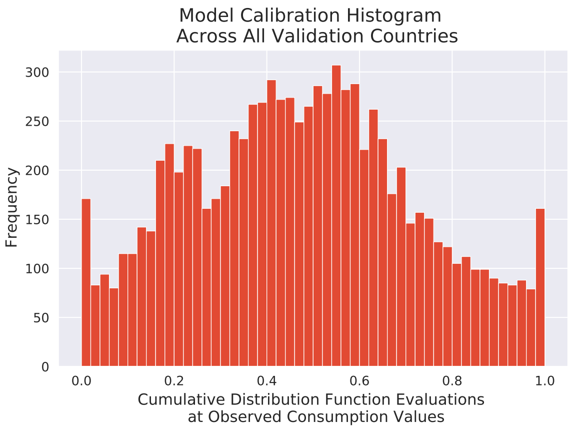


Figure 3: **Model calibration plot showing fit to data.** The 50-bin histogram represents the empirical frequency that a specific 2% probability interval was observed in a held-out set. Figure from [17] with permission.

used, and how specific outputs must be constrained to be consistent with other components of broader energy systems being modeled.

While we do not argue that inductive supervised machine learning methods should be seen as substitutive for deductive simulation models such as the WEM or WEPS, we believe that they offer decision-makers a valuable alternative and are made even more useful because they employ a complementary, data-driven approach to forecasting. They are reproducible and may engender additional trust, they have setups that are transparent and simple to explain, and are largely immune to negative aspects of human bias.

4.1.2. Model calibration

Another useful attribute of probabilistic demand forecasts is their ability to facilitate model calibration analyses. These analyses show how well forecasts made on held-out (unseen) data performed historically, and give an indication for how well one may expect the model to perform in the future.

Fig. 3 represents a model calibration histogram with 50 bins that shows the frequency that empirical electricity consumption fell into a 2% cumulative distribution function (CDF) intervals within predicted (posterior) probability distributions. These CDF intervals should have equal probability of occurrence. Because the probabilistic forecast follows the (non-uniform) negative binomial distribution, the intervals will generally correspond to different ranges in units of consumption. One way to think about these intervals is to consider each as representing one side of a 50-sided dice. Ideally, the dice should be fair: all sides are ideally equally probable, and with enough rolls, they should have more-or-less equal frequencies of occurrence. This is analogous to our histogram: a perfectly calibrated model would show a more-or-less uniform histogram. The non-uniform histogram in Fig. 3 suggests that the LDF model described in [17] is not perfectly calibrated. Analogously, the dice is not perfectly fair. Histograms closer to uniform distributions show evidence of better model calibration than those that are further away.

4.1.3. Determining feature importance and enabling improved interpretability

Some methods used for probabilistic electricity demand forecasting, including the LDF framework used in [17], facilitate the assessment of feature importance and provide insight into model behavior. These attributes can constitute novel domain-specific insight, in addition to comprise Value of Information (VoI) metrics that can be used to inform information investments, as discussed in Section 2.

In the LDF model, one way to describe feature importance is by determining the gradients of interpretable quantities such as the forecast mean and forecast standard deviation with respect to input features. A positive *forecast mean gradient* with respect to a given feature means that, as that feature is increased, forecasted *consumption* will increase on net. A positive forecast *standard deviation gradient* with respect to a given feature implies that as that feature is increased, forecast *uncertainty* will increase. A feature is more important for an

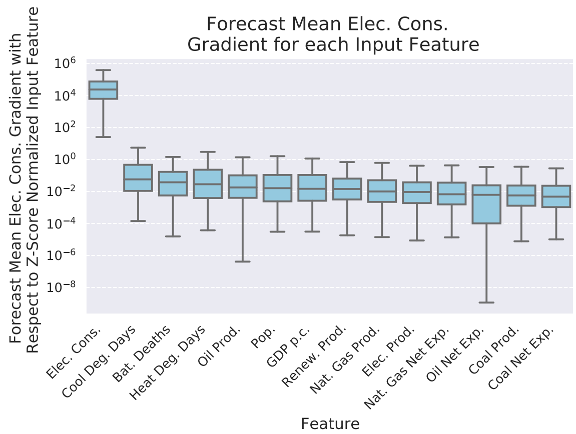


Figure 4: **Feature importance shown as the absolute value of forecast mean (top) and standard deviation gradient (bottom) across features.** Features are ranked in descending order of median importance. Figure from [17] with permission.

inference if model output is more sensitive to its value: this corresponds to a higher absolute value of normalized forecast mean or absolute value of forecast standard deviation gradient.

Fig. 4 depicts a box plot for the absolute value of forecast mean gradients relative to each input feature. From this analysis, it is evident that expected future electricity consumption is most affected by historical electricity consumption, followed in descending order by cooling degree days, battle-related deaths, heating degree days, oil production, population, GDP per capita, renewable energy production, natural gas production, electricity production, natural gas net exports, oil net exports, coal production, and coal net exports.

4.2. Limitations of machine learning-based probabilistic forecasts

Three notable limitations exist when using inductive probabilistic forecasting models: they can only be fit when sufficient training data is available, they are not robust to unexpected changes in the relationship between historical inputs and forecasts, and they may not be interpretable in ways that more deductive models can be.

4.2.1. More training data may be required

As described earlier, supervised machine learning models and their variations are *inductive*: patterns are identified from numerous observations of input-output data pairs. In the case of the LDF probabilistic forecasting models introduced earlier, algorithms are used to learn ground-up relationships between future consumption values and inputs including historical consumption, GDP per capita, population, cooling degree days, heating degree days, and other features to define calibrated output probability distributions. Because of the potential complexity of the relationships between time series input and output, hundreds of input-output pairings or more are likely required to learn a reliable model using available information. In contrast, *deductive* models (or models that are at least *more* deductive) often require less training data; relationships between historical data and input features can be defined by modelers using existing economic, social, or technical theory. While supervised machine learning models are optimized to explicitly fit historical data, they have comparatively higher data costs.

4.2.2. Models are not robust to unexpected changes in relationships between historical data and forecasts

The act of forecasting is inherently extrapolative; general relationships are learned from historical data to extrapolate into the future. Challengingly, relationships bewteen past and future may change over time. For instance, improvements to air conditioning efficiency may be thought to decrease the influence that cooling degree days (a measure of how much cooling is needed over a particular time span) has on electricity consumption.

Probabilistic forecasting models can capture *expected* changes to relationships between historical data and forecasts through characterizations of forecast uncertainty. If historical trends in these relationships are found to be unstable, the model will be rewarded if it increases forecast uncertainty. This is related to the assumption that future forecasting uncertainty should be similar to past forecasting uncertainty. This assumption falls apart, however, if sudden shifts in historical data-forecast relationships occur in ways and with magnitudes that are not observed in historical training data.

Following our previous example, a sudden and unexpected improvement in electric appliance efficiency would likely yield over-estimations of future electricity demand. Similar effects may be observed with changes to electricity prices, shifts in economic activities, consumer preferences, and other phenomena. While these limitations are common to all inductive forecasting models, other, more deductive frameworks may deal with them better. For instance, it may be easy for an expert to know that electricity prices will decrease at some point in the future given the construction of new local generating capacity. Incorporating this knowledge in deductive simulation-based models may be more straightforward than in models following inductive frameworks.

4.2.3. Models are still not fully interpretable

In Section 4.1.3 we describe how probabilistic forecasting models may lend themselves to analyses that afford improved levels of model interpretability. Specifically, feature-specific gradients from the LDF model are described that can inform analysts about how mean consumption values and forecast uncertainty change with input perturbations. Nevertheless, important aspects of models based on artificial neural networks such as the LSTM used in this application of the LDF model, remain difficult to interpret due to their complexity. In contrast, simpler models and models based on more deductive principles may be easier to interpret, and in some ways better garner human trust.

5. Conclusion

In this paper, we discuss ways in which probabilistic electricity demand forecasts can provide value to infrastructure planning endeavors in low-access countries. Among other things, they can enable investments that could not occur otherwise. By doing so, they have the potential to expand the resource pool available for electrification efforts by incentivizing private investment.

We also show evidence that such probabilistic methods are underrepresented in the literature when it comes to low-access countries. Additionally, even the most prominent point forecasts demonstrate high variability when compared to historical electricity consumption. We use these observations to highlight the need for probabilistic methods to be applied in this space.

Finally, we highlight a specific model used for probabilistic electricity demand forecasting: the LDF model. This model is based on methods for Bayesian inference and artificial neural networks and follows an inductive epistemology. We qualitatively describe useful attributes and limitations of the LDF model and similar promising methods for probabilistic forecasting. We do this with the hope of outlining key modeling concepts that decision-makers should know before employing such probabilistic forecasts.

6. Acknowledgements

Support for this research was provided by the Applied Research Programme on Energy and Economic Growth (EEG), the Enel Foundation, and the Energy for Growth Hub. We would specifically like to thank Claudio Pregagnoli, Giuseppe Montesano, and Benjamin Klooss for their thoughtful comments and reviews of paper drafts.

References

[1] United Nations, "Sustainable Development Goal 7," 2017. `https://sustainabledevelopment.un.org/sdg7`.

[2] IEA, *World Energy Outlook 2020*. 2020.

[3] International Energy Agency, "More people have access to electricity than ever before, but the world is falling short of its sustainable energy goals," 2019. `https://www.iea.org/newsroom/news/2019/may/sustainable-development-goal-7-tracking-report.html`.

[4] E. G. Cazalet, C. E. Clark, and T. W. Keelin, "Costs and benefits of over/under capacity in electric power system planning. final report, october 1978," tech. rep., Decision Focus, Inc., Palo Alto, CA (USA), 1978.

[5] A. A. Mir, M. Alghassab, K. Ullah, Z. A. Khan, Y. Lu, and M. Imran, "A review of electricity demand forecasting in low and middle income countries: The demand determinants and horizons," *Sustainability*, vol. 12, no. 15, p. 5931, 2020.

[6] E. Thomson, "Power shortages in china: why?," *China: An International Journal*, vol. 3, no. 1, pp. 155–171, 2005.

[7] S. Yu, L. Zhou, and C. Li, "China wrestles with power shortages," *POWER Magazine*, 2013.

[8] M. Dyson and A. Engel, "The billion-dollar costs of forecasting electricity demand," *Rocky Mountain Institute*, 2017.

[9] R. T. Kivaisi, "Installation and use of a 3 kwp pv plant at umbuji village in zanzibar," *Renewable energy*, vol. 19, no. 3, pp. 457–472, 2000.

[10] H. Louie and P. Dauenhauer, "Effects of load estimation error on small-scale off-grid photovoltaic system design, cost and reliability," *Energy for Sustainable Development*, vol. 34, pp. 30–43, 2016.

[11] S. Mandelli, C. Brivio, E. Colombo, and M. Merlo, "Effect of load profile uncertainty on the optimum sizing of off-grid pv systems for rural electrification," *Sustainable Energy Technologies and Assessments*, vol. 18, pp. 34–47, 2016.

[12] N. Moksnes, A. Korkovelos, D. Mentis, and M. Howells, "Electrification pathways for kenya–linking spatial electrification analysis and medium to long term energy planning," *Environmental Research Letters*, vol. 12, no. 9, p. 095008, 2017.

[13] F. Riva, F. Gardumi, A. Tognollo, and E. Colombo, "Soft-linking energy demand and optimisation models for local long-term electricity planning: An application to rural india," *Energy*, vol. 166, pp. 32–46, 2019.

[14] R. Amatya, M. Barbar, Y. Borofsky, M. Brusnahan, P. Ciller, T. Cotterman, F. de Cuadra, C. Drouin, P. Duenas, D. Ellman, A. González-García, S. J. Lee, V. Li, C. Mateo, O. Oladeji, R. Palacios, I. Pérez-Arriaga, R. Stoner, and C. Vergara, "Computer-aided electrification planning in developing countries: The Reference Electrification Model (REM)," *IIT Working Paper*, 2018.

[15] P. Ciller, D. Ellman, C. Vergara, A. González-García, S. J. Lee, C. Drouin, M. Brusnahan, Y. Borofsky, C. Mateo, R. Amatya, *et al.*, "Optimal electrification planning incorporating on-and off-grid technologies: the reference electrification model (rem)," *Proceedings of the IEEE*, vol. 107, no. 9, pp. 1872–1905, 2019.

[16] S. J. Lee, E. Sánchez Jacob, A. González García, P. Ciller Cutillas, P. Dueñas Martínez, J. Taneja, F. Cuadra García, J. Lumbreras Martín, H. Daly, R. J. Stoner, and I. J. Pérez-Arriaga, "Investigating the necessity of demand characterization and stimulation for geospatial electrification planning in developing countries," *MIT Center for Energy and Environmental Policy Research*, 2019.

[17] S. J. Lee, C. L. Dean, D. Suri, P. Somani, J. Pacheco, R. Stoner, I. Perez-Arriaga, J. W. Fisher III, and J. Taneja, "Probabilistic forecasts of country-level electricity demand in Africa," *arxiv*, 2021.

[18] H. H. Gossen, *The laws of human relations and the rules of human action derived therefrom*. MIT Press (MA), 1983.

[19] M. J. Kochenderfer, *Decision making under uncertainty: theory and application*. MIT press, 2015.

[20] I. J. Perez-Arriaga, R. Stoner, R. Rahnama, S. J. Lee, G. Jacquot, E. Protzer, A. Gonzalez, R. Amatya, M. Brusnahan, P. Duenas, *et al.*, "A utility approach to accelerate universal electricity access in less developed countries: A regulatory proposal,"

[21] I. J. Perez-Arriaga, R. J. Stoner, D. Nagpal, and G. Jacquot, "Global commission to end energy poverty: Inception report," *MIT Energy Initiative*, 2019.

[22] R. De Neufville and S. Scholtes, *Flexibility in engineering design*. MIT Press, 2011.

[23] C. B. Agaton and H. Karl, "A real options approach to renewable electricity generation in the philippines," *Energy, Sustainability and Society*, vol. 8, no. 1, p. 1, 2018.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory, Second Edition*. 2006.

[25] S. J. Lee, "Adaptive electricity access planning," 2018.

[26] C. Brown, P. Alexander, A. Arneth, I. Holman, and M. Rounsevell, "Achievement of paris climate goals unlikely due to time

lags in the land system," *Nature Climate Change*, vol. 9, no. 3, pp. 203–208, 2019.
[27] O. Hoegh-Guldberg, D. Jacob, M. Taylor, T. G. Bolaños, M. Bindi, S. Brown, I. A. Camilloni, A. Diedhiou, R. Djalante, K. Ebi, *et al.*, "The human imperative of stabilizing global climate change at 1.5 c," *Science*, vol. 365, no. 6459, 2019.
[28] IPCC, "Mitigation of climate change," *Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, vol. 1454, 2014.
[29] C. E. Casillas and D. M. Kammen, "The energy-poverty-climate nexus," *Science*, vol. 330, no. 6008, pp. 1181–1182, 2010.
[30] J. D. Jenkins, M. Luke, and S. Thernstrom, "Getting to zero carbon emissions in the electric power sector," *Joule*, vol. 2, no. 12, pp. 2498–2510, 2018.
[31] W. Boyd, "Public utility and the low-carbon future," *UCLA L. Rev.*, vol. 61, p. 1614, 2013.
[32] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," 2016.
[33] C. L. Dean, S. J. Lee, J. Pacheco, and J. W. Fisher III, "Lightweight data fusion via conjugate mappings," *arXiv preprint arXiv:2011.10607*, 2020.
[34] IEA, *World Energy Outlook 1998*. 1998.
[35] IEA, *World Energy Outlook 2000*. 2000.
[36] IEA, *World Energy Outlook 2002*. 2002.
[37] IEA, *World Energy Outlook 2004*. 2004.
[38] IEA, *World Energy Outlook 2006*. 2006.
[39] IEA, *World Energy Outlook 2008*. 2008.
[40] IEA, *World Energy Outlook 2009*. 2009.
[41] IEA, *World Energy Outlook 2010*. 2010.
[42] IEA, *World Energy Outlook 2011*. 2011.
[43] IEA, *World Energy Outlook 2012*. 2012.
[44] IEA, *World Energy Outlook 2013*. 2013.
[45] IEA, *World Energy Outlook 2014*. 2014.
[46] IEA, *World Energy Outlook 2015*. 2015.
[47] IEA, *World Energy Outlook 2016*. 2016.
[48] IEA, *World Energy Outlook 2017*. 2017.
[49] IEA, *World Energy Outlook 2018*. 2018.
[50] IEA, *World Energy Outlook 2019*. 2019.
[51] US EIA, *International Energy Outlook 1995*. 1995.
[52] US EIA, *International Energy Outlook 1996*. 1996.
[53] US EIA, *International Energy Outlook 1997*. 1997.
[54] US EIA, *International Energy Outlook 1998*. 1998.
[55] US EIA, *International Energy Outlook 1999*. 1999.
[56] US EIA, *International Energy Outlook 2000*. 2000.
[57] US EIA, *International Energy Outlook 2001*. 2001.
[58] US EIA, *International Energy Outlook 2002*. 2002.
[59] US EIA, *International Energy Outlook 2003*. 2003.
[60] US EIA, *International Energy Outlook 2004*. 2004.
[61] US EIA, *International Energy Outlook 2005*. 2005.
[62] US EIA, *International Energy Outlook 2006*. 2006.
[63] US EIA, *International Energy Outlook 2007*. 2007.
[64] US EIA, *International Energy Outlook 2008*. 2008.
[65] US EIA, *International Energy Outlook 2009*. 2009.
[66] US EIA, *International Energy Outlook 2010*. 2010.
[67] US EIA, *International Energy Outlook 2011*. 2011.
[68] US EIA, *International Energy Outlook 2013*. 2013.
[69] US EIA, *International Energy Outlook 2016*. 2016.
[70] US EIA, *International Energy Outlook 2017*. 2017.
[71] US EIA, *International Energy Outlook 2019*. 2019.
[72] H. Liao, J.-W. Cai, D.-W. Yang, and Y.-M. Wei, "Why did the historical energy forecasting succeed or fail? a case study on iea's projection," *Technological Forecasting and Social Change*, vol. 107, pp. 90–96, 2016.
[73] IEA, "World energy model documentation, 2020 version," 2020.
[74] O. of Energy Analysis, "World energy projection system model documentation," tech. rep., U.S. Energy Information Administration, 2011.
[75] C. Wolfram, "What the heck is happening in the developing world?," *Energy Institute Blog*, 2016.
[76] L. Börjeson, M. Höjer, K.-H. Dreborg, T. Ekvall, and G. Finnveden, "Scenario types and techniques: towards a user's guide," *Futures*, vol. 38, no. 7, pp. 723–739, 2006.
[77] S. S. Adebola, "Electricity consumption and economic growth: Trivariate investigation in botswana with capital formation," *International Journal of Energy Economics and Policy*, vol. 1, no. 2, pp. 32–46, 2011.
[78] A. Adedokun, "Nigeria electricity forecast and vision 20: 2020: Evidence from arima model," *Energy Sources, Part B: Economics, Planning, and Policy*, vol. 11, no. 11, pp. 1027–1034, 2016.
[79] O. Adeoye and C. Spataru, "Modelling and forecasting hourly electricity demand in west african countries," *Applied Energy*, vol. 242, pp. 311–333, 2019.
[80] P. K. Adom and W. Bekoe, "Conditional dynamic forecast of electrical energy consumption requirements in ghana by 2020: A comparison of ardl and pam," *Energy*, vol. 44, no. 1, pp. 367–380, 2012.
[81] M. Bazilian, P. Nussbaumer, H.-H. Rogner, A. Brew-Hammond, V. Foster, S. Pachauri, E. Williams, M. Howells, P. Niyongabo, L. Musaba, *et al.*, "Energy access scenarios to 2030 for the power sector in sub-saharan africa," *Utilities Policy*, vol. 20, no. 1, pp. 1–16, 2012.
[82] BP, *Energy Outlook 2020*. 2020.
[83] R. Van Buskirk, "Analysis of long-range clean energy investment scenarios for eritrea, east africa," *Energy policy*, vol. 34, no. 14, pp. 1807–1817, 2006.
[84] D. Chikobvu and C. Sigauke, "Regression-sarima modelling of daily peak electricity demand in south africa," *Journal of Energy in Southern Africa*, vol. 23, no. 3, pp. 23–30, 2012.
[85] ExxonMobil, *Outlook for Energy: A perspective to 2040*. 2019.
[86] O. Ezennaya, O. Isaac, U. Okolie, and O. Ezeanyim, "Analysis of nigerias national electricity demand forecast (2013–2030)," *International Journal of Science and Technology Research*, vol. 3, no. 3, pp. 333–340, 2014.
[87] I. A. Ezenugu, S. C. Nwokonko, and I. Markson, "Modelling and forecasting of residential electricity consumption in nigeria using multiple and quadratic regression models," *American Journal of Software Engineering and Applications*, vol. 6, no. 3, pp. 99–104, 2017.
[88] S. Guefano, J. G. TAMBA, T. E. W. Azong, and L. Monkam, "Forecast of electricity consumption in the cameroonian residential sector by grey and vector autoregressive models," *Energy*, p. 118791, 2020.
[89] R. Inglesi, "Aggregate electricity demand in south africa: Conditional forecasts to 2030," *Applied energy*, vol. 87, no. 1, pp. 197–204, 2010.
[90] IEA, *World Energy Outlook 2019*. 2019.
[91] A. Miketa and B. Merven, "West african power pool: Planning and prospects for renewable energy," *IRENA, Abu Dhabi*, 2013.
[92] IRENA, "Global renewables outlook: energy transformation 2050 (edition: 2020)," 2020.
[93] M. E. Lebotsa, C. Sigauke, A. Bere, R. Fildes, and J. E. Boylan, "Short term electricity demand forecasting using partially linear additive quantile regression with an application to the unit commitment problem," *Applied Energy*, vol. 222, pp. 104–118, 2018.
[94] L. Marwala and B. Twala, "Forecasting electricity consumption in south africa: Arma, neural networks and neuro-fuzzy systems," in *2014 International Joint Conference on Neural Networks (IJCNN)*, pp. 3049–3055, IEEE, 2014.

[95] P. Mokilane, J. Galpin, V. Sarma Yadavalli, P. Debba, R. Koen, and S. Sibiya, "Density forecasting for long-term electricity demand in south africa using quantile regression," *South African Journal of Economic and Management Sciences*, vol. 21, no. 1, pp. 1–14, 2018.

[96] A. S. Momodu, A. Addo, J.-F. K. Akinbami, and Y. Mulugetta, "Low-carbon development strategy for the west african electricity system: preliminary assessment using system dynamics approach," *Energy, Sustainability and Society*, vol. 7, no. 1, p. 11, 2017.

[97] G. Okoboi and J. Mawejje, "Electricity peak demand in uganda: insights and foresight," *Energy, Sustainability and Society*, vol. 6, no. 1, p. 29, 2016.

[98] N. S. Ouedraogo, "Africa energy future: Alternative scenarios and their implications for sustainable development strategies," *Energy Policy*, vol. 106, pp. 457–471, 2017.

[99] N. S. Ouedraogo, "Modeling sustainable long-term electricity supply-demand in africa," *Applied energy*, vol. 190, pp. 1047–1067, 2017.

[100] E. Panos, H. Turton, M. Densing, and K. Volkart, "Powering the growth of sub-saharan africa: the jazz and symphony scenarios of world energy council," *Energy for Sustainable Development*, vol. 26, pp. 14–33, 2015.

[101] E. Panos, M. Densing, and K. Volkart, "Access to electricity in the world energy council's global energy scenarios: An outlook for developing regions until 2030," *Energy Strategy Reviews*, vol. 9, pp. 28–49, 2016.

[102] S. A. Sarkodie, "Estimating ghanas electricity consumption by 2030: An arima forecast," *Energy Sources, Part B: Economics, Planning, and Policy*, vol. 12, no. 10, pp. 936–944, 2017.

[103] K. Shibano and G. Mogi, "Electricity consumption forecast model using household income: Case study in tanzania," *Energies*, vol. 13, no. 10, p. 2497, 2020.

[104] C. Sigauke, "Forecasting medium-term electricity demand in a south african electric power supply system," *Journal of Energy in Southern Africa*, vol. 28, no. 4, pp. 54–67, 2017.

[105] C. Sigauke and A. Bere, "Modelling non-stationary time series using a peaks over threshold distribution with time varying covariates and threshold: An application to peak electricity demand," *Energy*, vol. 119, pp. 152–166, 2017.

[106] C. Sigauke and D. Chikobvu, "Daily peak electricity load forecasting in south africa using a multivariate non-parametric regression approach," *ORiON*, vol. 26, no. 2, 2010.

[107] R. Spalding-Fecher, M. Senatla, F. Yamba, B. Lukwesa, G. Himunzowa, C. Heaps, A. Chapman, G. Mahumane, B. Tembo, and I. Nyambe, "Electricity supply and demand scenarios for the southern african power pool," *Energy Policy*, vol. 101, pp. 403–414, 2017.

[108] C. Taliotis, A. Shivakumar, E. Ramos, M. Howells, D. Mentis, V. Sridharan, O. Broad, and L. Mofor, "An indicative analysis of investment opportunities in the african electricity supply sectorusing temba (the electricity model base for africa)," *Energy for Sustainable Development*, vol. 31, pp. 50–66, 2016.

[109] L. Tartibu and K. Kabengele, "Forecasting net energy consumption of south africa using artificial neural network," in *2018 International Conference on the Industrial and Commercial Use of Energy (ICUE)*, pp. 1–7, IEEE, 2018.

[110] United States Energy Information Administration, *International Energy Outlook 2020*. 2020.

[111] World Energy Council, *World Energy Scenarios: composing energy futures to 2050*. 2013.

Supplemental Information

6.1. IEA and EIA Historical Forecasts for India, China, and the U.S.

Fig. 5, 6, and 7 show historical electricity demand and historical forecasts from the IEA and EIA for India, China, and the U.S., respectively.

6.2. Predictive, Explorative, and Normative Models

Future energy scenarios can be classified according to the types of questions they seek to answer. *Predictive scenarios* answer the question "what will happen in the future?" Predictive scenarios are often informed by data on the historical and current states of a system being analyzed. *Explorative scenatios* seek to answer the question, "what can happen in the future?", and generally outline points along a range of possible conditional outcomes. Lastly, *normative scenarios* answer the question "how can a specific target be reached?" These scenarios are often constructed by modeling some future state of a system, and then working backwards to understand how those outcomes could be realized [76].

Each of these three modeling types have applications that they are well-suited for; however, predictive scenarios are most useful for planning concrete investments in electricity infrastructure in low-access countries. Planners need to consider forecasts of what is actually likely to occur in the future to right-size network, generation, and storage investments.

Table 1 shows a non-exhaustive review of existing electricity forecasting studies for countries and regions in Africa. In these tables, we classify models by predictive, explorative, and normative model types; whether or not the model is probabilistic; the number of national-level forecasts presented; whether or not forecasts are provided at the region- or continent-level; whether or not forecasts are global in scope; and the forecast time horizon. The LDF model [17] described in Section 4 is highlighted in the table.

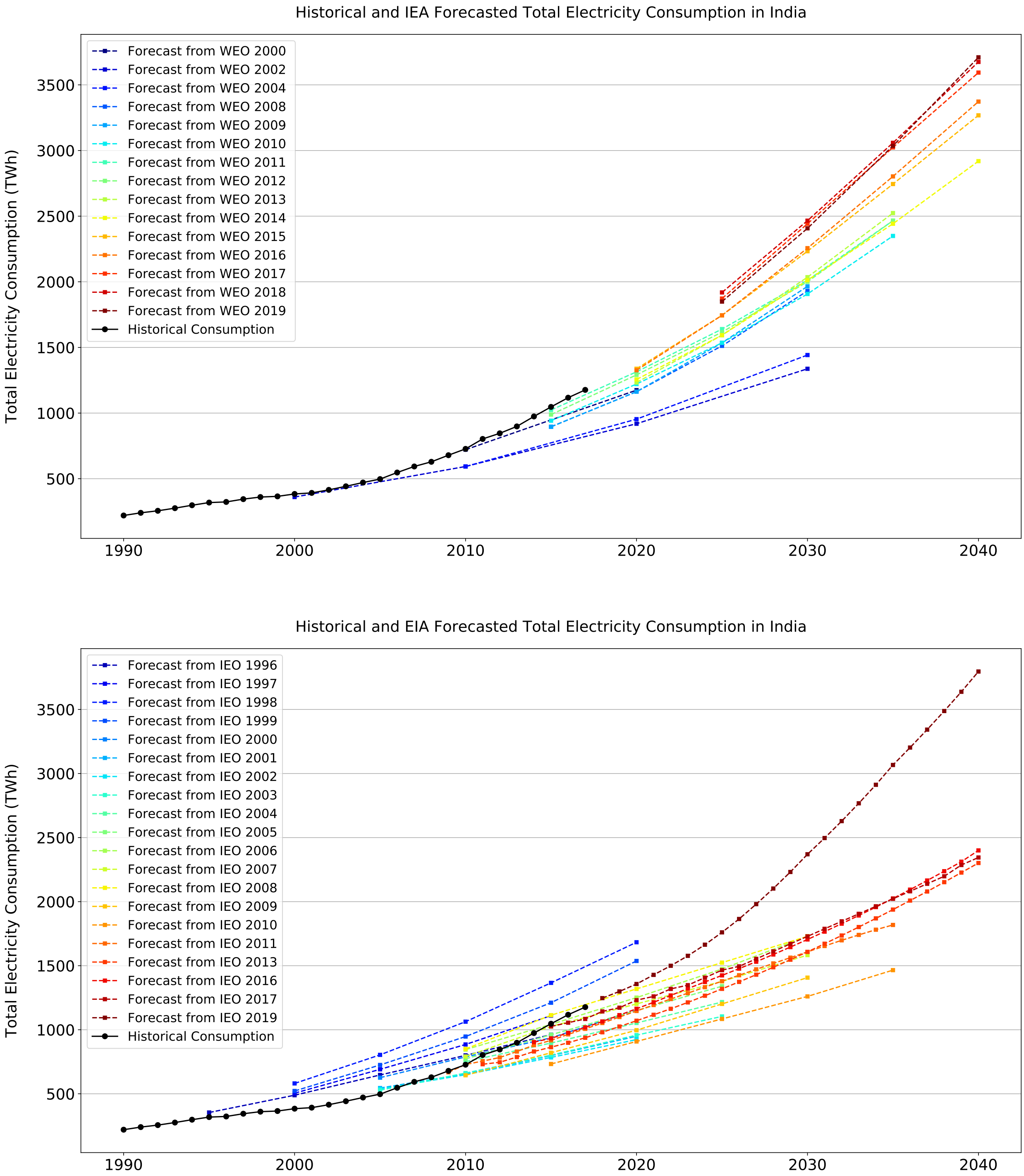
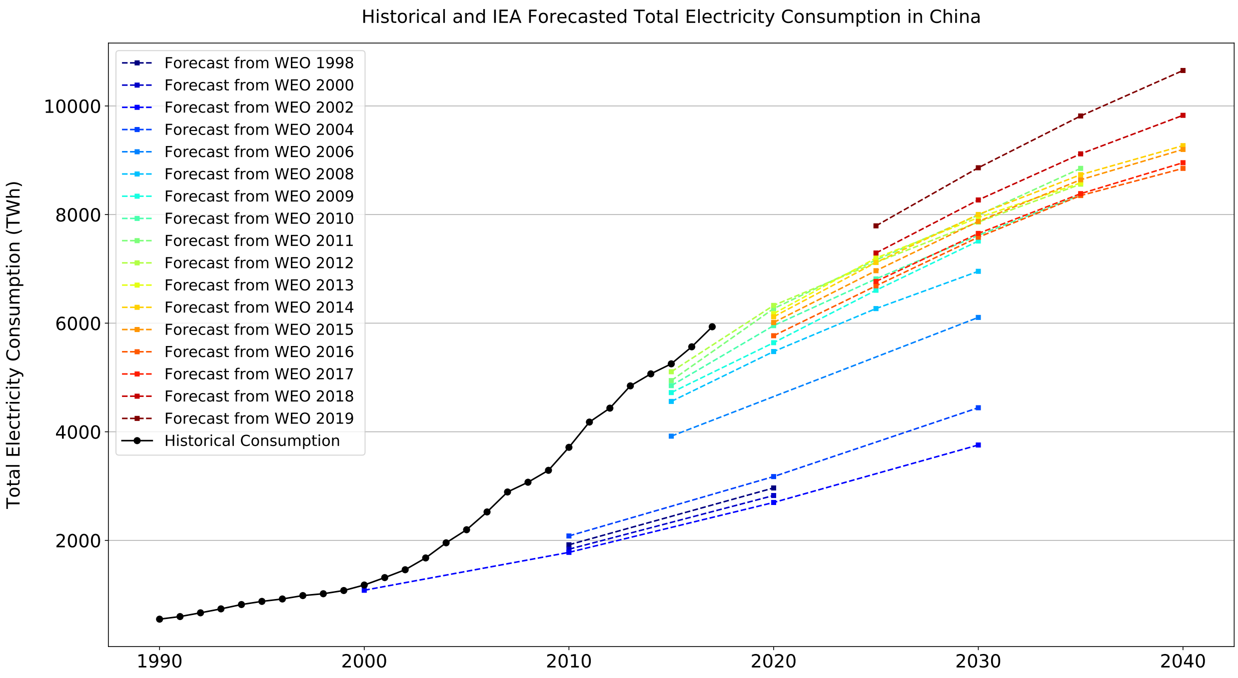


Figure 5: **Comparing historical electricity demand and historical forecasts for India**. Forecasts are shown from the IEA (top) and EIA (bottom).



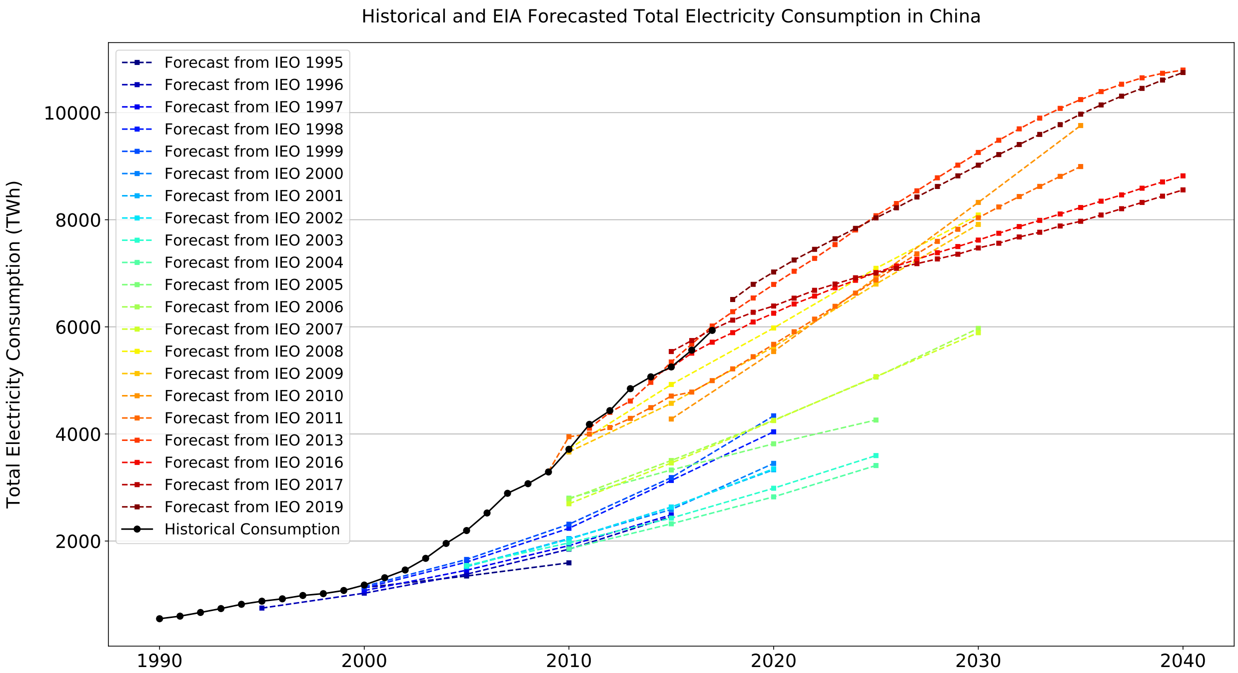
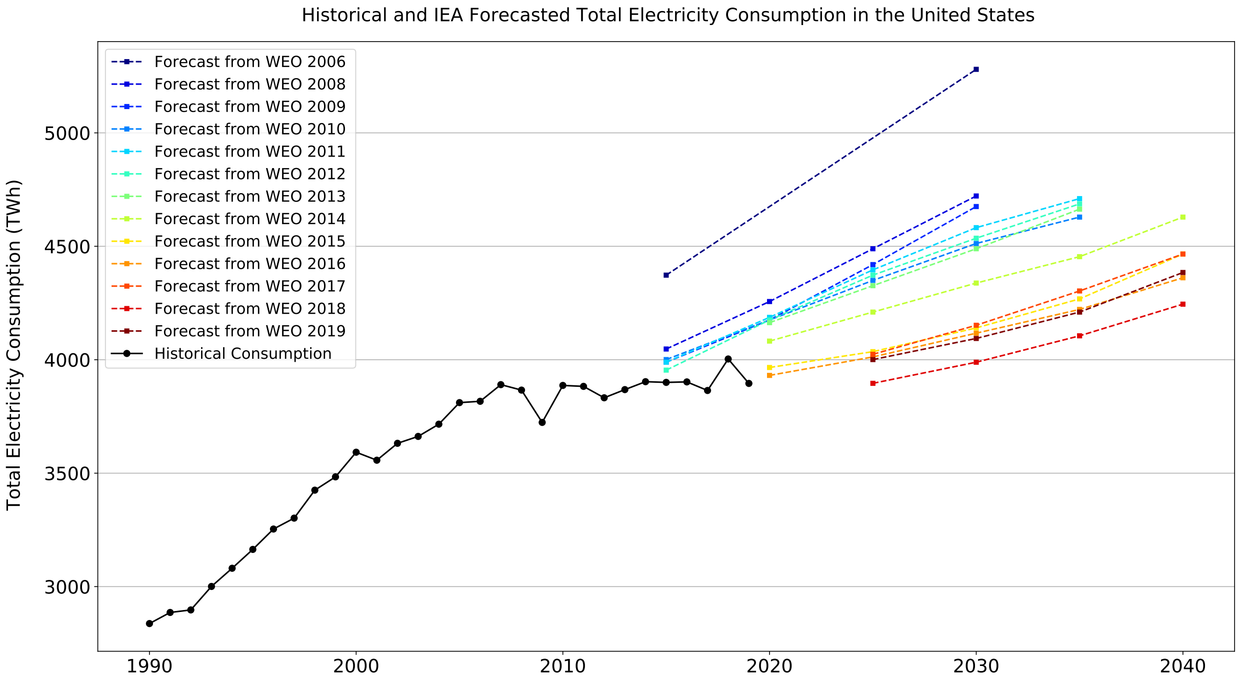


Figure 6: **Comparing historical electricity demand and historical forecasts for China**. Forecasts are shown from the IEA (top) and EIA (bottom).



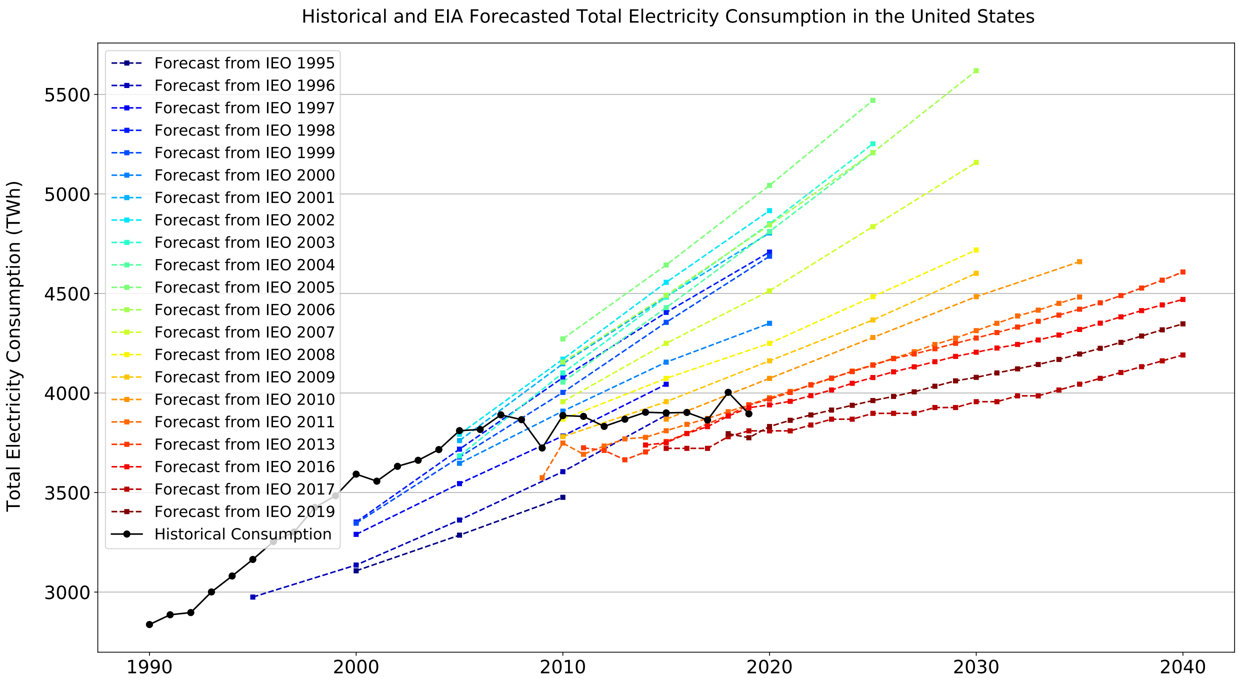


Figure 7: **Comparing historical electricity demand and historical forecasts for the United States**. Forecasts are shown from the IEA (top) and EIA (bottom).

Table 1: Table with attributes of electricity demand forecasting studies in Africa.

	Author	Year	Title	Methodology	Model Type				Geographic Scope			Forecast Time Horizon	Ref.
					Predictive	Explorative	Normative	Probabilistic	National-level	Reg. or Cont.-level	Global		
1	Lee et al.	2020	Probabilistic forecasts of country-level electricity demand in Africa	LDF, LSTM	✓			✓	52	✓		2016-2031	[17]
2	Adebola	2011	Electricity consumption and economic growth: Trivariate investigation in Botswana with capital formation	ARIMA	✓				1			2009-2016	[77]
3	Adedokun	2016	Nigeria electricity forecast and vision 2020: Evidence from ARIMA model	ARIMA	✓	✓	✓		1			2012-2050	[78]
4	Adeoye and Spataru	2019	Modelling and forecasting hourly electricity demand in West African countries	Bottom-up; Linear Model-Based	✓				14			2017-2030	[79]
5	Adom and Bekoe	2012	Conditional dynamic forecast of electrical energy consumption requirements in Ghana by 2020: A comparison of ARDL and PAM	ARDL; PAM	✓				1			2009-2020	[80]
6	Bazilian et al.	2012	Energy access scenarios to 2030 for the power sector in sub-Saharan Africa	Heuristic		✓	✓			✓		2010-2030	[81]
7	BP	2020	Energy Outlook 2020	Proprietary		✓				✓	✓	2018-2050	[82]
8	Van Buskirk	2006	Analysis of long-range clean energy investment scenarios for Eritrea, East Africa	Economic, Deductive	✓	✓			1			2000-2100	[83]
9	Chikobvu and Sigauke	2012	Regression-SARIMA modelling of daily peak electricity demand in South Africa	SARIMA; Regression-SARIMA	✓			✓	1			2009 (14 days)	[84]
10	ExxonMobil	2019	Outlook for Energy: A perspective to 2040	Proprietary	✓		✓			✓	✓	2017-2040	[85]
11	Ezennaya	2014	Analysis of Nigeria's national electricity demand forecast (2013-2030)	Linear Model-Based	✓				1			2013-2030	[86]
12	Ezenugu et al.	2017	Modelling and Forecasting of residential electricity consumption in Nigeria using Multiple and Quadratic regression models	Multiple and Quadratic Regression	✓				1			2015-2029	[87]
13	Guefano et al.	2020	Forecast of electricity consumption in the Cameroonian residential sector by Grey and vector autoregressive models	Grey and Vector Autoregressive Models	✓				1			2018-2025	[88]
14	Inglesi	2010	Aggregate electricity demand in South Africa: Conditional forecasts to 2030	Linear Model-Based		✓			1			2006-2030	[89]

	Author	Year	Title	Methodology	Model Type: Predictive	Model Type: Explorative	Model Type: Normative	Probabilistic	Geographic Scope: National-level	Geographic Scope: Reg. or Cont.-level	Geographic Scope: Global	Forecast Time Horizon	Ref.
15	International Energy Agency	2020	World Energy Outlook	Econometric Methods; Simulation-Based Model	✓	✓	✓		1	✓	✓	2025-2040	[90]
16	International Renewable Energy Agency	2013	West African power pool: Planning and prospects for renewable energy	Extrapolation	✓				13			2010-2050	[91]
17	International Renewable Energy Agency	2020	Global Renewables Outlook: Energy transformation 2050	Bottom-up		✓	✓			✓	✓	2016-2050	[92]
18	Lebotsa et al.	2018	Short term electricity demand forecasting using partially linear additive quantile regression with an application to the unit commitment problem	Partially Linear Additive Quantile Regression	✓			✓	1			2012 (hourly)	[93]
19	Marwala and Twala	2014	Forecasting electricity consumption in South Africa: ARMA, neural networks and neuro-fuzzy systems	ANNs; ARMA	✓				1			2012-2023	[94]
20	Mokilane et al.	2018	Density forecasting for long-term electricity demand in South Africa using quantile regression	Quantile Regression	✓			✓	1			2013-2023	[95]
21	Momodu et al.	2017	Low-carbon development strategy for the West African electricity system: preliminary assessment using System dynamics approach	Systems Dynamics		✓				✓		2015-2060	[96]
22	Okoboi and Mawejje	2016	Electricity peak demand in Uganda: insights and foresight	Linear Model-Based (Double Exponential Forecasting Model)	✓				1			2014-2021	[97]
23	Ouedraogo	2017	Africa energy future: Alternative scenarios and their implications for sustainable development strategies	Bottom-up (LEAP)		✓				✓		2010-2040	[98]
24	Ouedraogo	2017	Modeling sustainable long-term electricity supply-demand in Africa	Bottom-up (LEAP)		✓				✓		2015-2040	[99]
25	Panos et al.	2015	Powering the growth of Sub-Saharan Africa: the jazz and symphony scenarios of World Energy Council	Bottom-up (MARKAL)		✓				✓		2050	[100]

					Model Type				Geographic Scope				
	Author	Year	Title	Methodology	Predictive	Explorative	Normative	Probabilistic	National-level	Reg. or Cont.-level	Global	Forecast Time Horizon	Ref.
26	Panos et al.	2016	Access to electricity in the World Energy Council's global energy scenarios: An outlook for developing regions until 2030	Bottom-up (MARKAL)		✓	✓			✓		2010, 2020, 2030	[101]
27	Sarkodie	2017	Estimating Ghana's electricity consumption by 2030: An ARIMA forecast	ARIMA	✓				1			2014-2030	[102]
28	Shibano and Mogi	2020	Electricity Consumption Forecast Model Using Household Income: Case Study in Tanzania	Bottom-up	✓				1			2010	[103]
29	Sigauke	2017	Forecasting medium-term electricity demand in a South African electric power supply system	GAMs	✓				1			2013	[104]
30	Sigauke and Bere	2017	Modelling non-stationary time series using a peaks over threshold distribution with time varying covariates and threshold: An application to peak electricity demand	Generalized Pareto Distribution Model	✓				1			2000-2010	[105]
31	Sigauke and Chikobvu	2010	Daily peak electricity load forecasting in South Africa using a multivariate non-parametric regression approach	Multivariate Non-parametric Regression	✓				1			2009 (44 days)	[106]
32	Spalding-Fecher et al.	2017	Electricity supply and demand scenarios for the Southern African power pool	Bottom-up		✓			12			2015-2070	[107]
33	Taliotis et al.	2016	An indicative analysis of investment opportunities in the African electricity supply sector Using TEMBA (The Electricity Model Base for Africa)	Bottom-up (TEMBA)		✓			45			2020-2040	[108]
34	Tartibu and Kabengele	2018	Forecasting net energy consumption of South Africa using artificial neural network	ANN		✓			1			2014-2050	[109]
35	U.S. Energy Information Administration	2020	International Energy Outlook 2020	Econometric-Methods; Simulation-Based Model	✓	✓				✓	✓	2018-2050	[110]
36	World Energy Council	2013	World Energy Scenarios: composing energy futures to 2050	Bottom-up (MARKAL)		✓	✓			✓	✓	2020-2050	[111]