RESEARCH SUMMARY

System Utilization Metrics

Operational Efficiency & Asset Utilization: A Research Synthesis

Prepared by Planmetrix LLC

1. Purpose & Context

This summary consolidates a broad research effort exploring how electric distribution utilities and their regulators measure operational efficiency and asset utilization. The report proceeds in stages: an initial survey of North American practice centered on load factor, a broadening to international and cross-industry comparisons, a synopsis of promising next-generation metrics, and a focused case study of Data Envelopment Analysis (DEA) as implemented by Finland's electricity regulator.

The motivating question is a practical one. Load factor — the ratio of average load to peak load — is the metric most utilities and regulators reach for first when discussing asset utilization, largely because it is simple and familiar. But across more than 30 case studies spanning North America, Europe, Africa, Australasia, and South America, the research finds load factor to be a useful diagnostic and a poor regulatory instrument: it cannot distinguish a utility making smart infrastructure investments from one simply experiencing a mild winter, and it actively penalizes some energy-efficiency and distributed-generation programs that regulators want to encourage.

What an Ideal Metric Should Do

Before evaluating any specific metric, the report lays out four criteria that a well-designed efficiency metric should satisfy. It should reward deferred capital investment, better infrastructure utilization, and diverse distributed generation; it should not penalize efficiency programs, electric vehicles, or conservative reliability design choices; it should be insensitive to weather, load composition, and the basic demographic character of a service territory; and it should be based on data that is readily available, understandable, trackable year over year, and comparable across regions. Every metric examined is ultimately judged against this checklist — and, as the synthesis below shows, none satisfies it fully on its own.

A note on scope: the underlying research is explicitly framed as a literature review and synthesis — surveying existing regulatory practice, academic studies, and utility disclosures — rather than as original primary research or as a finished metric specification. Its value lies in the breadth of the comparison it assembles in one place: more than 30 jurisdictions and case studies, cross-referenced against a consistent advantages/disadvantages framework, which is precisely the kind of comparative groundwork that later metric-design work can build on.

2. Load Factor: The Starting-Point Metric

Load factor is defined as the ratio of average load to peak (maximum) load over a given period. The appeal is its simplicity: a utility with a low load factor is paying to build and maintain generation, transmission, and distribution assets that sit idle most of the time. Research cited in the report suggests that as little as 1% of annual hours can account for 10–18% of total capacity requirements in North America — a striking illustration of how much infrastructure exists solely to serve a handful of peak hours per year.

Both directly and indirectly, a range of levers can move load factor: energy-efficiency programs, time-of-use and dynamic pricing, demand response, electric vehicle charging behavior, and smart grid investment all interact with it, sometimes in counterintuitive ways. A program that reduces air-conditioning load, for instance, improves load factor because it cuts peak more than average consumption; a program that reduces street lighting load tends to worsen it, because lighting runs mostly off-peak. One cited 2010 FERC survey found that many respondent utilities saw load reductions equivalent to 30% of peak load through demand response — yet nearly half of all load-reduction programs in 2001 were never actually called upon, suggesting substantial untapped headroom in existing demand-response capability.

Advantages and Disadvantages

Advantages	Disadvantages
Scalable across an entire system or individual customer classes; versatile enough to describe generation, transmission, or distribution efficiency alike.	Does not correlate well with the specific projects that relieve capacity constraints, and gives no insight into generation mix, operating conditions, or customer demographics.
Simple enough to explain to non-specialists and useful for basic energy audits at the individual-customer level.	Penalizes distributed generation, since intermittent solar and wind reduce average load without reliably reducing peak.
Can approximate other engineering metrics and serves as a reasonable proxy for how well peak demand is managed in broad terms.	Highly sensitive to weather and to underlying changes in load composition that a utility cannot control.
Long track record and wide familiarity make it easy to communicate to boards, regulators, and the public.	Too imprecise for rate-setting: a utility that cuts peak load by 80 MW (5%) conveys far more than a 2-point change in load factor, and comparisons across utilities are unreliable.

Key Finding

In a 2013 sample of 11 New York ISO regions, four utilities posted load factors above 60% and four fell below 50% — a spread wide enough to make cross-utility comparison nearly meaningless without substantial normalization. The report concludes that load factor is a reasonable rule-of-thumb diagnostic but is “too crude to be anything other than” that at the regulatory level.

How Utilities Actually Use It: Selected Case Studies

The report’s appendix surveys roughly 20 North American case studies, and a few illustrate the range of approaches well. Fort Pierce Utilities Authority in Florida offers large customers (billable demand over 40 kW) a literal load-factor discount: a credit ranging from $0.50 to $2.00 per kW of billed demand, scaled to load factors between 60% and 75%-plus. PacifiCorp, operating across six western states, instead treats energy efficiency as a competing supply-side resource in its planning models rather than as a simple load modifier — a framing that let efficiency programs outcompete over 2,500 MW of potential new generation on cost grounds. Xcel Energy in Colorado committed to DSM programs projected to deliver $1.3 billion in net customer benefits, while Nova Scotia Power paired time-of-use rates with electric thermal storage incentives that grew peak-demand savings from 5.7 MW in 2008 to 50.7 MW by 2012.

Other metrics reviewed alongside load factor in the report include capacity factor (output versus rated capacity rather than peak load, often used for renewables) and availability factor (the share of time a resource can generate, 80–99% for gas turbines). Capacity factor has its own blind spot, though: consider a spot network with four 1,000-kVA transformers feeding a commercial building, each averaging 300 kW of load. Measured against nameplate rating, that is a 30% capacity factor — but if the system is designed to N-2 reliability criteria (able to carry the full load even with two transformers out of service), the real usable capacity is 2,000 kVA, and the true capacity factor is 60%. Ignoring reliability design this way can make a well-engineered, appropriately redundant system look twice as wasteful as it actually is.

The Regulatory Machinery Behind These Numbers

Most of the U.S. examples above trace back to Integrated Resource Planning (IRP), a framework dating to the 1992 Energy Policy Act that requires utilities to evaluate the full range of supply-side and demand-side alternatives before committing to new capacity. States implement IRP unevenly: Missouri and Washington treat energy efficiency as a supply-side resource competing directly against generation, while Maine, California, and Massachusetts instead treat it as a priority resource to be exhausted before any new capacity is considered. FERC has layered federal policy on top of this state-level patchwork in four steps — the 1992 and 2005 Energy Policy Acts, which respectively opened wholesale markets to independent producers and made demand response official federal policy, followed by 2008’s Order 719 and 2011’s Order 745, which together require regional grid operators to price demand response at full market value for comparable products.

Because energy-efficiency and demand-response programs reduce the very sales utilities are otherwise compensated for, regulators pair IRP with financial mechanisms designed to keep utility incentives aligned with customer and system benefits — rate decoupling, performance-based incentive bonuses for exceeding commission-mandated goals, and system benefit charges that set aside a share of revenue specifically for efficiency and demand-response investment. Without one of these mechanisms in place, a utility doing exactly what regulators ask — cutting its own sales — has little financial reason to keep doing it.

Five Ways to Test Whether a Program Is Worth It

Once a program exists, its cost-effectiveness is judged through one or more of five standard benefit-cost tests, each representing a different stakeholder's point of view. The Participant Cost Test (PCT) asks whether the customers actually enrolling come out ahead. The Ratepayer Impact Measure (RIM) — originally called the Non-Participant Test — asks whether the program raises average prices for everyone else; a RIM benefit-cost ratio below 1.0 signals it will. The Program Administrator Cost Test (PACT) takes the utility's own viewpoint, comparing avoided supply costs to program costs. The Total Resource Cost Test (TRC) takes the broadest service-territory view, weighing total avoided supply costs against total program costs; 18 of 27 U.S. states surveyed rely on the TRC or its variant as their sole test. The Societal Cost Test (SCT) goes one step further, adding non-energy benefits like avoided pollution that fall outside the utility's service territory altogether.

No single test is sufficient on its own: the PACT treats efficiency the same as any other supply resource and is arguably the cleanest standalone test, while the RIM can be used to block efficiency investment entirely even when it passes every other test. The report recommends running the TRC and SCT together as the primary efficiency screen, then layering on PCT, PACT, and RIM results to show regulators how costs and benefits land differently across customer groups — since a program that passes the TRC but fails the RIM will still raise costs for the customers who never participate in it.

The Department of Energy's Smart Grid Research and Development Multi-Year Program Plan illustrates how ambitious these targets can get on paper: it set a goal, to be reached by 2030, of a 40% improvement in system efficiency and asset utilization sufficient to lift load factor to 70%, alongside 20% of electricity capacity (200 GW) from distributed and renewable sources. The report notes a recurring theme across nearly every jurisdiction surveyed: the chief obstacle to tracking progress against targets like these is not a lack of ambition but a lack of comparable, quantitative benchmarking data — the same gap that motivates the international and benchmarking research that follows.

3. International & Cross-Industry Perspectives

This section asks whether other countries, or other infrastructure industries entirely, have solved the asset-utilization measurement problem more elegantly. The short answer is no single country has, but several have built more sophisticated regulatory machinery around the underlying challenge than North America has, and other industries reveal genuinely different ways of framing the same trade-off between idle capacity and service quality.

Country Practices

Country / Region	Notable Practice	Interesting Detail
United Kingdom	RIIO framework (Revenue = Incentives + Innovation + Outputs) succeeded the older RPI-X model; the electric distribution iteration, RIIO-ED1, covers 2015–2023.	Ofgem's Load Index sorts substations into five loading bands (L1–L5) but is applied only to primary and grid substations, not lower-voltage assets.
New Zealand	29 Electricity Lines Businesses use load factor internally, but it is not a regulator-mandated metric.	Orion estimates that holding 1990's 50.7% load factor would have required 750 MW of peak capacity instead of 630 MW — an estimated NZD 18 million per year in avoided distribution and transmission cost.
South Africa	SANEDI has set an explicit national target: a 40% asset-utilization	Eskom's FLEXICON/COMRICON water-heater control pilot (9,000

Country / Region	Notable Practice	Interesting Detail
	improvement to reach a 70% load factor.	households, 1999) aimed to cut a 33 MW winter peak by 10 MW.
Australia	ElectraNet uses load factor to benchmark against other National Electricity Market regions.	South Australia has the lowest load factor in the NEM, and rising air-conditioning penetration is cited as the main driver of decline nationwide.

Lessons from Other Industries

Natural gas: is structurally easier to balance because it can be stored cheaply. The U.S. operates more than 100 peak-shaving LNG plants that liquefy and bank gas for release during cold spells, and average pipeline utilization across the 30 largest U.S. suppliers was 59% in 2007 — with five providers exceeding 90%. Because storage absorbs most of the peak-versus-average mismatch, load-factor-style metrics matter far less to gas regulators than to electric ones.

Telecommunications: frames the same problem around quality of service rather than utilization efficiency. Operators often deliberately build excess capacity to protect latency and throughput, particularly for real-time applications like VoIP, and one striking data point from an Australian broadband study found that the residential "edge" of the network runs at just 1.7% of capacity on average — underutilization that exists specifically to absorb each evening's viewing-and-gaming peak.

Water utilities: tend to over-build deliberately, prioritizing public-health margin and future growth capacity over near-term utilization efficiency, though rising price pressure is gradually pushing more conservative forecasting.

Key Finding

Outside North America, load factor is used in only a handful of markets (New Zealand, Australia, South Africa) and almost always as an internal management metric rather than a regulatory one. The clearer international trend — especially across the EU — is a shift toward benchmarking frameworks that evaluate a utility's overall performance rather than isolating any single ratio.

China: A Cautionary Growth Story

China's experience illustrates what happens when demand growth outpaces the regulatory tools meant to manage it. Between 1993 and 2004, residential and commercial energy use grew 17% and 12% respectively, and the country was projected to need an additional 160 GW of capacity by 2020. Air-conditioning-driven peak growth pushed load factors down even as a high share of coal-fired generation — which cannot easily ramp down off-peak — made the decline costlier to absorb. Beijing's experiment with time-of-use pricing, charging peak rates three to four times the off-peak rate, held load factor roughly flat between 1997 and 2003 despite an otherwise prevailing 5% decline, but the approach has not scaled nationally, partly because subsidized retail rates elsewhere blunt the price signal that makes TOU effective.

Ontario: Modeling the Cost of Flexibility

Ontario's grid illustrates the economics behind every peak-shaving argument in this report. The province's base load of roughly 13,000 MW is met mostly by nuclear and hydroelectric generation — cheap to run but slow to ramp — while peak loads up to 27,000 MW require oil, natural gas, and flexible hydro capacity that costs substantially more to operate. Demand was projected to climb to 175 TWh by 2025 at roughly 0.5% annual growth, driven by population growth and heating/cooling demand. The province's own Canadian Integrated Modeling System (CIMS) simulation found that demand-response and dynamic-pricing programs could shave up to 10% of peak demand, and the Ontario Power Authority subsequently set a formal target of 6,300 MW in demand reduction by 2025 — with internal estimates suggesting almost double that level might be achievable given better coordination between government agencies, public education, and long-term efficiency funding.

Gas Industry Detail: Why Storage Changes Everything

The gas sector's comparative advantage over electricity is mechanical, not regulatory: natural gas can be liquefied, stored, and re-gasified on demand, while electricity at grid scale largely cannot. Gas distribution peaks are driven primarily by heating demand and tracked through Heating Degree Days (the gap between 65°F and the daily average temperature) rather than the daily peaks that define electric system planning. Pipelines are required to satisfy an extreme winter peak with roughly a 1-in-20-year statistical probability, and EU Directive 2009/73/EC requires member states to file annual reports on supply-demand balance, planned capacity additions, and shortfall contingencies — a transparency requirement with no direct electric-sector equivalent in the research surveyed.

4. Benchmarking as the Emerging Consensus

Benchmarking compares utilities against each other, or against a hypothetical "ideal" company, rather than tracking any one utility against a fixed target. It does not produce a single number so much as a distribution — the best-performing utilities define a "frontier," and regulators set the acceptable performance bar somewhere along that frontier, often slightly below it to leave room for measurement error and to preserve some financial upside for utilities that improve.

Two Families, Four Techniques

Frontier benchmarking compares each utility to the single most efficient peer; it requires large, high-quality, comparable datasets but rewards the most aggressive convergence toward best practice. Average benchmarking instead compares each utility to a group mean and is more forgiving of smaller or noisier datasets, at the cost of less precision. Within those two families, the literature identifies four broad statistical techniques.

Technique	Approach	Where It's Used
Parametric (OLS, SFA)	Statistical cost-function estimation with explicit assumptions about the underlying distribution.	DEA and SFA together are the most common methods; used by regulators in the UK, Switzerland, and Scandinavia, among others
Non-parametric (DEA)	Linear-programming approach with no distributional assumptions; handles multiple inputs/outputs simultaneously.	Same UK/Switzerland/Scandinavia base, plus Germany's 198-company study and Finland's EMA model
Hybrid (SDEA)	Combines parametric and non-parametric elements to balance the strengths of each.	Selected EU regulators
Engineering models	Builds an artificial "ideal" company from expert judgment rather than observed data.	Used where comparable peer data is scarce

Selected Benchmarking Results

- **United Kingdom:** average DEA-based asset-utilization score of 77.7% across utilities.
- **Netherlands:** 73.1%, broadly comparable to the UK figure.
- **Slovenia:** 32.6%, a result the researchers flag as indicating substantial room for improvement.
- **Germany:** a 198-company DEA/SFA study found an average score of 92.2%, ranging from 75.5% to 100%.
- **Iran:** a DEA/PCA study found 14 of the companies sampled sitting exactly on the efficiency frontier, with an average score of 78% and roughly two-thirds of companies needing to reduce inputs to catch up.

Weather consistently complicates these comparisons. A U.S. transmission study covering 2001–2009 found that once weather controls were added, asset utilization had actually declined slightly over the period — the opposite conclusion from an unadjusted reading. A separate South American study spanning Argentina, Brazil, Chile, and Peru incorporated data from 429 meteorological stations and 3,423 NASA lightning-strike coordinates, and found that while weather did not shift the average utility's score much, it meaningfully reordered the rankings for individual outlier utilities — exactly the companies most likely to dispute a benchmarking-based rate decision.

Additional International Case Studies

Norway: applied SFA across 118 distribution utilities from 2004 to 2012, using a Malmquist productivity index to separate efficiency gains from broader technical change. Productivity improved measurably after the country switched to incentive-based regulation, with the largest gains concentrated among previously low-performing utilities — consistent with the general finding that incentive regulation does more for laggards than for utilities already near the frontier.

Colombia: offers a cautionary counterpoint. Following major blackouts in 1992 and 1993, market reforms delivered clear reliability gains, but a dynamic stochastic frontier study found efficiency improvements were wildly uneven across companies, with small rural utilities improving fastest. Customers did not see meaningful quality-of-service improvement until roughly a decade after the reforms began, and prices rose alongside the efficiency gains — a reminder that a rising benchmarking score and a satisfied customer base are not the same thing.

European gas transmission: benchmarking, studied by the Electricity Policy Research Group using DEA, COLS, and SFA across 40 U.S. and 4 European operators, found that — with one exception — the European operators were consistently less efficient than their U.S. counterparts. The study's own recommendation for building on the result was to push for better data standardization across European regulators, since comparisons against U.S. utilities are useful but a more standardized European dataset would make any future benchmarking-based tariff decisions more defensible.

5. Additional Regulator Case Studies

Beyond the headline DEA scores cited above, the report's appendix compiles a dozen narrower regulator studies that illustrate how widely DEA and SFA methodology choices vary even within the same general approach. A selection of the more instructive examples:

Jurisdiction	Method	Notable Result
New Zealand	Total Factor Productivity (TFP), 1996–2003	Productivity rose 3% per year from 1996–2000 and 2.5% per year from 2001–2003, averaging 2% annually, driven mainly by falling operating costs across distribution companies.
Brazil	DEA + Bayesian SFA, 60 utilities	Combining both techniques reduced information asymmetry and addressed outlier sensitivity, a known weakness of standalone deterministic DEA models.
Germany (structure study)	DEA with constant returns to scale, SFA for verification	Customer density mattered significantly only in the bottom third of performers; grid composition (cable vs. aerial) had no measurable effect, and peak load proved less important than base load.
Iran	DEA + Principal Component Analysis, bootstrap confidence intervals	PCA reduced nine candidate inputs to four (customer density, service area, energy loss rate, OpEx) feeding two outputs (customers served, energy supplied).
Germany (DEA/SFA benchmarking)	Four parallel models: DEA I/II, SFA I/II across 198 companies	Running standardized-capital-cost variants alongside company-reported-cost variants let the regulator cross-check whether capital reporting differences were distorting results.

Taken together, these studies reinforce a point worth making explicitly: there is no single correct DEA specification. Regulators in different countries deliberately choose different inputs, outputs, and environmental controls to fit their own utility populations, and the choice of method matters less than the discipline of testing results against a second technique (SFA, Bayesian methods, or bootstrapping) to confirm the ranking is not an artifact of one particular model.

Who Is Actually Using Benchmarking

The regulators identified as active benchmarking users span every continent represented in the case studies: Ofgem in the UK, the Commerce Commission in New Zealand, the Ontario Energy Board in Canada, Ireland's CER, and the full range of Australian state regulators, alongside a long tail of academic studies that regulators draw on before formalizing their own methodology. A 40-country survey of energy regulators across Europe, Australasia, and Latin America found benchmarking widespread in principle but inconsistent in execution: EU countries tend to cluster around similar techniques (with process/activity methods and DEA the dominant choices for electricity overall), while Australian and Latin American regulators cite limited regional data and limited in-house benchmarking experience as the main reasons they have not adopted frontier methods more fully.

6. Promising Next-Generation Metrics

This section steps back from the historical survey and proposes four specific metrics worth further development, organized into two categories: asset utilization (normalized peak load, substation demand factor) and economic efficiency ($/kWh delivered, DEA). Each is evaluated against the same advantages/disadvantages framework applied to load factor earlier in this report.

Asset Utilization Metrics

Metric	Key Advantage	Key Disadvantage	Development Needed
Normalized Peak Load	Encourages demand reduction, storage, and TOU rates without penalizing renewables or efficiency programs.	Doesn’t directly indicate asset condition or where investment is actually needed.	Weather normalization methods; possible customer-count adjustment.
Substation Demand Factor	Ties directly to the infrastructure that drives capital spending, unlike system-wide load factor.	Hard to scale from individual substations to a system-wide figure; both very high and very low factors signal problems.	A scaling formula (the paper proposes one, summing the share of substations above 90% and below 50% loading).

Economic Efficiency Metrics

Metric	Key Advantage	Key Disadvantage	Development Needed
$/kWh Delivered	Simple, intuitive, and encourages real cost-benefit discipline across programs.	May penalize end-use solar and efficiency programs; sensitive to storms and major infrastructure events.	Decide whether to include OPEX only, CAPEX only, or both, and how to adjust for major events.
Data Envelopment Analysis	Folds multiple cost drivers into a single comparative efficiency score; widely used by European regulators already.	Harder to explain to non-specialists; results depend heavily on which inputs and outputs are chosen.	Requires a large, comparable utility sample and dedicated operations-research expertise to formulate properly.

Key Finding

The report’s proposed system-level substation-scaling formula is a genuinely novel contribution: $K_{system} = P_{overloaded} + P_{underloaded}$, where $P_{overloaded}$ is the share of substations exceeding 90% of allowable demand and $P_{underloaded}$ is the share below 50%. A utility could, in principle, drive this number toward zero through right-sized investment rather than uniform peak shaving — addressing one of load factor’s core blind spots directly.

For DEA specifically, the report lists representative inputs (operational expenses, capital expenses) and outputs (number of customers, overhead and underground line lengths, kWh delivered, peak load) that a utility-comparison model might use, while cautioning that the appropriate peer set for a large urban utility like Con Edison would likely differ from the peer set appropriate for a smaller utility like NYSEG — reinforcing that sample design, not just technique selection, drives result quality.

7. Finland Case Study: DEA in Practice

This section examines the one DEA implementation treated as a fully worked example: Finland's Energy Market Authority (EMA), which has applied some form of DEA-based benchmarking since 1998. A 2000 project, run jointly with the Helsinki School of Economics and industry representatives, re-examined the model across Finland's roughly 106 distribution companies — a set varied enough to span dense urban networks and sparse rural ones within a single regulatory framework.

A separate case study elsewhere in this report adds useful context on why Finland is a good test market in the first place: it describes Finland as one of the most competitive electricity markets in the EU, evidenced by very high rates of customer switching, with 88 distribution companies setting their own tariffs (a slightly different count than the ~106 companies cited in the EMA's 2000 re-examination, likely reflecting market consolidation or a different reference year) within methods set by the EMA, including Capex and Opex calculations for rate-of-return purposes. That case study describes a model using both DEA and SFA, with inputs covering controllable operating costs, depreciation, and outage costs, and outputs covering network length, customer count, and value of electricity consumed — a usefully different lens from the dedicated DEA deep-dive below, which focuses on the EMA's specific input/output selection process rather than the full cost-of-service model.

Inputs, Outputs, and Environmental Controls

The dedicated DEA re-examination project's model is more tightly specified than the broader cost-of-service framework above, built around a small number of carefully chosen variables:

Category	Variable	Measurement Basis
Input	Operational Expenditure (OpEx)	Euros, with uncontrollable costs (e.g., wholesale electricity prices) subtracted out. Capital expenditure was considered and rejected due to year-to-year volatility.
Output	Distributed Energy	Financial value, weighted by national voltage-level-based distribution prices across three voltage tiers.
Output	Service Quality	Moving average of total customer interruption time.
Environmental	Number of Customers	Used to normalize for differences in customer density across utilities.
Environmental	Network Length (3 voltage levels)	Kilometers; a proxy for geographic dispersion and the higher costs it implies.
Environmental	Average Winter Snow Depth	Centimeters; a proxy for weather-driven outage risk.
Environmental	Forest Cover	Square kilometers; a proxy for rural construction cost and interruption risk.

Two design choices stand out. First, capital expenditure was deliberately excluded as an input precisely because it is so volatile year to year — a pragmatic acknowledgment that a technically "complete" model can be less useful than a stable one. Second, the environmental variables (snow depth, forest cover, customer geography) are not efficiency measures at all; they exist purely to keep the comparison fair between a dense Helsinki-area utility and a sparse rural one operating under genuinely different physical conditions.

How the Score Feeds Into Rates

The DEA efficiency score is not, by itself, the regulatory outcome. It is one input into a larger calculation of each utility's "realized adjusted profit" — starting from accounting profit, then adjusting for costs the regulator does not recognize and for a basket of incentive mechanisms covering efficiency, investment, quality, and innovation. Over a four-year regulatory period, the cumulative surplus or deficit against a target return is carried forward: utilities that ran a surplus must return it to customers in the following period, while utilities that ran a deficit are permitted to recover it. The structure deliberately keeps regulatory incentives tied to each company's actual accounting performance rather than treating the efficiency score as a one-time pass/fail test.

This four-year carry-forward mechanism is itself worth noting as a design choice other regulators could adapt. By summing surpluses and deficits across the full regulatory period rather than settling accounts annually, the Finnish model gives utilities room to absorb a single bad year — a major storm, an unusual cost spike — without an immediate rate-case fight, while still ensuring that, over the full period, the company's allowed return tracks its actual efficiency performance rather than drifting away from it.

Key Finding

Finland's model demonstrates that a workable DEA framework needs only a small number of carefully chosen inputs and outputs — one input (OpEx) and two outputs (energy delivered, service quality) — provided the environmental controls are well chosen. The complexity lives in the normalization, not the core efficiency calculation.

8. Conclusions

Read together, the research converges on a consistent position. Load factor is a useful, simple, and widely understood diagnostic for getting a rough read on system performance, but it is not precise enough, and not resistant enough to weather and demographic noise, to serve as a standalone regulatory metric anywhere it has been seriously tested. The international survey confirms this is not a uniquely North American problem — even markets that use load factor more deliberately, such as New Zealand and South Africa, treat it as one input among several rather than a sole rate-setting criterion.

The clearer path forward, validated across the UK, Germany, Finland, and a growing share of EU regulators, is benchmarking — particularly DEA and SFA approaches that evaluate a utility's overall cost efficiency against peers rather than isolating any single ratio. These methods are more data-intensive and harder to explain to a general audience, but they account for the genuine differences in operating environment (weather, density, geography) that make naive utility-to-utility comparison unreliable, and they are already operating at meaningful scale: Finland's 106-company implementation and Germany's 198-company study are both functioning, ongoing regulatory programs rather than academic exercises.

The research also identifies a credible middle path for utilities not ready to adopt full DEA benchmarking: normalized peak load and substation demand factor, alongside a more disciplined $/kWh delivered metric, offer meaningful improvements on load factor's blind spots while remaining far simpler to calculate, explain, and audit than a full frontier-analysis program.

The research closes with a forward-looking observation worth carrying into any future metric design: load factor, capacity factor, and other metrics discussed in this report are built around average load, and average load is a steadily less meaningful concept as the grid fills with distributed generation, storage, and microgrids. A community that covers nearly all of its own consumption with local solar and storage does not need much net energy from the utility, but it still imposes real, irregular inflows and outflows that the utility must be able to handle at any moment. Billing and performance metrics built around kilowatt-hours delivered will increasingly miss the point if the utility's real job is shifting from supplying energy to managing the timing, location, and two-way flow of it.

9. Recommended Next Steps

The research points toward a staged approach: validate the simpler, lower-risk metrics first, while building the data infrastructure that a future DEA or SFA benchmarking program would require.

#	Action	Rationale
1	Pilot substation demand factor at a handful of constrained substations.	Directly ties a metric to the assets that actually drive capital spending, addressing load factor's weakest point.
2	Develop and test a weather-normalization method for normalized peak load.	Industry-accepted normalization approaches already exist; this is described as the lowest-risk next step in the research.
3	Define the system-level substation scaling formula	Converts a substation-level metric into something reportable at the system or regulatory level.

#	Action	Rationale
	(K_system) and test it against historical loading data.	
4	Inventory available OpEx, customer, network-length, and weather data against the Finnish DEA model's variable list.	Identifies data gaps before committing to a full benchmarking build-out.
5	Decide the appropriate peer set and sample size for any future DEA comparison.	Sample comparability is flagged as a harder problem than technique selection itself.

None of these steps requires committing upfront to a full benchmarking program. Items 1 and 2 can be piloted with data most utilities already collect, and both produce metrics that are interpretable on their own even if a DEA program is never built. Items 3 through 5 are explicitly preparatory: they convert the lessons of the Finnish and German implementations into a concrete checklist for whether, and how, a comparable program could eventually be built domestically, without requiring a decision on that larger question today.

A Note on this Summary

This summary draws on roughly 235 footnoted sources across the underlying report, spanning academic journals, regulatory filings, utility integrated resource plans, and government reports from more than a dozen countries. Readers interested in a specific case study, statistical method, or regulatory citation referenced above should consult the full report, where the complete source list and underlying data tables are preserved. The material is best read in the sequence presented here: establishing load factor as the baseline and showing its limits, broadening the lens internationally and across industries to confirm those limits are not a North American artifact, proposing concrete next-generation alternatives, and finally grounding the most promising of those alternatives, DEA, in a real, currently operating regulatory implementation.