WHITE PAPER 5

From Scenario to Real Time

One Verifiable Risk Method from Grid Planning to Operation

Svendsen* · Nyiredy · Milenkovic****

November 2026 · Version 1.0 · November 2026

Contact: arne@thefuturepowergrid.org · robert@thefuturepowergrid.org · mathieu@thefuturepowergrid.org · info@thefuturepowergrid.org

Published by The Future Power Grid. Authors write in an independent capacity unless otherwise stated.

Suggested citation: Svendsen, A. B., Nyiredy, R., & Milenkovic, M. (2026). From Scenario to Real Time: One Verifiable Risk Method from Grid Planning to Operation. White Paper 5. The Future Power Grid. Version 1.0, November 2026.

"It is difficult to make predictions, especially about the future." The past, however, can be predicted again — and checked. That is what makes a risk method worth trusting.

This paper builds on: Svendsen (2026), Origins and Operational Record, White Paper 0; and Svendsen, Nyiredy & Milenkovic (2026), White Papers 1–3. It is a companion to White Paper 4 (forthcoming).

* *Lead author and corresponding author*

** *Contributing authors*

Table of Contents

Glossary of Acronyms 3
Source Attribution 3
Executive Summary 4
1. Context: Building on the Series 6
2. One Main Task and One Main Goal 7
3. Why Now: Regulation Is Moving to Risk 8
4. Today's Analyses Show Half the Picture 9
5. A Complete System Analysis in Four Steps 10
6. One Method Along the Whole Time Axis 12
7. Verifiability: The Future We Predict, the Past We Verify 13
8. A Precedent from Banking 15
9. A Bridge Between Grid Companies 16
10. From Promise to Contract: Closing the Loop 17
11. Caveats and Qualifying Conditions 18
12. Conclusions and Specific Recommendations 19
References 20

Glossary of Acronyms

Acronym	Definition
ACER	Agency for the Cooperation of Energy Regulators
CSAM	Methodology for coordinating operational security analysis (EU)
DSA	Dynamic Security Assessment
DSO	Distribution System Operator
EENS / ENS	Expected Energy Not Supplied / Energy Not Supplied
ENTSO-E	European Network of Transmission System Operators for Electricity
FRCR	Frequency Risk and Control Report (Great Britain)
IEA	International Energy Agency
N-1	Single-contingency operational criterion
N-X	Multi-contingency state (X elements out of service)
NESO	National Energy System Operator (Great Britain)
Ofgem	Office of Gas and Electricity Markets (Great Britain)
P&L	Profit and loss (banking)
PRA	Probabilistic Risk Assessment
SQSS	Security and Quality of Supply Standard (Great Britain)
TSO	Transmission System Operator
VaR	Value-at-Risk (banking)

Source Attribution

Throughout this paper, a distinction is made between:

- "Regulatory and policy sources" referring to requirements, recommendations and findings directly reported by ACER, the UK Government, the IEA and the Basel Committee.
- "Operational experience" referring to results observed in practical implementations, as documented in White Paper 0 (Svendsen, 2026).
- "Authors' interpretation" referring to the framework, the proposed method and the conclusions drawn from these sources.

The method proposed in this paper is a framework for verification, not a calibrated result. Where the paper describes zones, thresholds or validation layers for grid risk models, these are proposals that require system-specific design and testing before they are applied in planning, operation or regulation.

This paper forms the fifth in a series. White Paper 0 documents the operational record behind the series. White Papers 1–3 present the three-layer framework, Virtual N-1 and the qualification of large flexible loads. White Paper 4 (forthcoming) addresses price and contract for conditional connection. This paper addresses the question that connects them: how one risk method can be applied consistently from planning to operation, and verified against what actually happens.

Executive Summary

Every grid operator has the same main task and the same main goal. The task — keeping the system in balance within frequency and voltage limits — is deterministic: does the system withstand the event? The goal — the highest possible security of supply at the lowest overall cost — is probabilistic: how likely is the event, and what does it cost? Traditional power system analysis documents the task. Probabilistic Risk Assessment (PRA) documents the goal.

Today the two are separated along the value chain. Investment and connection studies use one set of tools and criteria; operation uses another. A promise made in a planning study cannot be checked against what happens in operation, because the two are not expressed in the same terms. This paper proposes a governance and verification framework for probabilistic grid operation: a way to make risk-based decisions auditable, from the planning study to the control room. It rests on four elements:

- **A complete system analysis has four steps.** Defining the cases, electrotechnical feasibility, regularity analysis (PRA) and economic analysis. The third step is the one not assessed satisfactorily today, and without it "operationally justifiable" cannot be quantified.
- **One method along the whole time axis.** The same goal, model basis and acceptance criterion from investment decisions decades ahead to system operation now. Only the uncertainty of the inputs changes: scenarios, then forecasts, then measurements.
- **Verifiability.** A PRA model can be checked against history by replaying a recorded operating year and comparing expected with observed disturbances, weather-driven risk and energy not supplied. Banking supervisors have required this kind of backtesting of risk models since 1996. The backtest proposed here has not yet been carried out for a whole grid; it should be piloted with at least one grid owner before it is used in regulation.
- **A bridge.** Applied consistently, the method gives TSOs, DSOs and regulators a common risk currency — and an audit trail from the risk accepted in planning to the risk observed in operation.

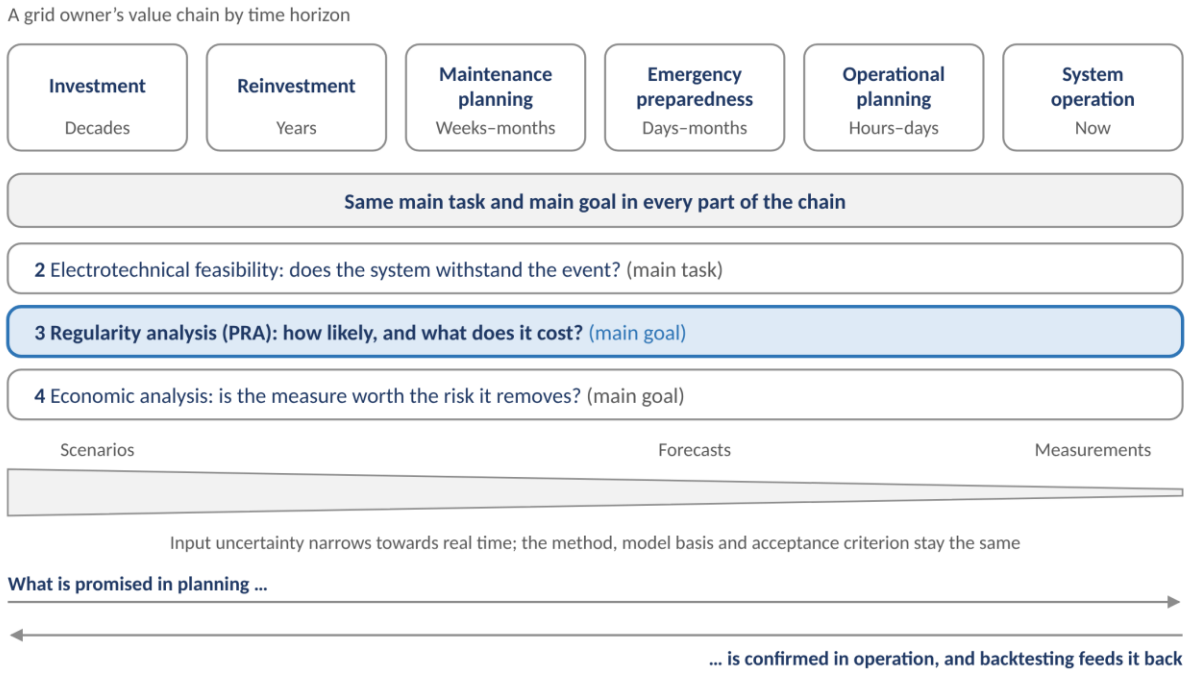


Figure 1 *One method from decades ahead to real time. Steps 2–4 are applied consistently in every part of the value chain; only the uncertainty of the inputs changes. Conceptual illustration.*

What is promised in planning can be confirmed in operation — but only if planning and operation speak the same risk language, and only if that language can be checked against what actually happened.

The timing is set by regulation. ACER's CSAM requires all European TSOs to develop a common methodology for probabilistic risk assessment in operational security analysis by 31 December 2027. The UK Government's independent review recommends full risk-based grid operation by 2035 and planning by 2036. In Great Britain, NESO and Ofgem already set frequency-security policy on a probabilistic cost–risk basis, and the regulator's response to its most recent proposal shows that what delays acceptance is validation, not the principle (Section 7.1). If operation becomes probabilistic by regulation while planning remains deterministic, the gap between them will widen. The elements of the method proposed here have been demonstrated in operation (White Paper 0). What remains is to make them consistent, systematic and verifiable.

1. Context: Building on the Series

The White Paper series has argued from several angles that the grid is the binding constraint on electrification, and that a large part of the answer lies in operating existing assets on the basis of quantified risk (IEA, 2026a; IEA, 2026b). White Paper 1 proposed a three-layer framework with PRA and DSA as the Decision Layer. White Paper 2 extended it with Virtual N-1. White Paper 3 applied it to large flexible loads such as data centres. White Paper 0 documented that the method behind the series has already been used in projects, in a national-scale prototype and in commercial operation.

These papers address what risk-based operation can do. They leave open a practical question that grid owners, regulators and study engineers raise as soon as the argument is accepted: how is the same risk assessment used consistently across planning and operation — and how do we know that it is right? White Paper 0 concludes that the remaining challenge is “adoption, not invention”. This paper argues that adoption depends on two things: one consistent method along the value chain, and a way to verify it.

This gives White Paper 5 a different role from the papers before it. White Papers 1–4 describe what risk-based operation can do and how it is implemented. This paper describes how it can be governed: how a grid owner, a regulator or a counterparty to a connection agreement can check that the method works. It is, in that sense, the governance and verification layer of the series (Figure 2).

Paper	Question answered	Contribution
White Paper 0	Why should the method be trusted?	25-year operational and project record
White Paper 1	How can existing grids carry more, safely?	Three-layer framework (PRA/DSA, GET, flexibility)
White Paper 2	What is N-1, and what comes after it?	Quantified N-1 and Virtual N-1
White Paper 3	Which large loads can become grid resources?	Qualification framework for flexible consumers
White Paper 4 (forthcoming)	How should conditional connection be priced and contracted?	Price and contract terms
White Paper 5 (this document)	How is one risk method used from planning to operation, and verified?	Governance and verification: four-step analysis, one method along the time axis, backtesting

Table 1 *The White Paper series and the question each paper answers*

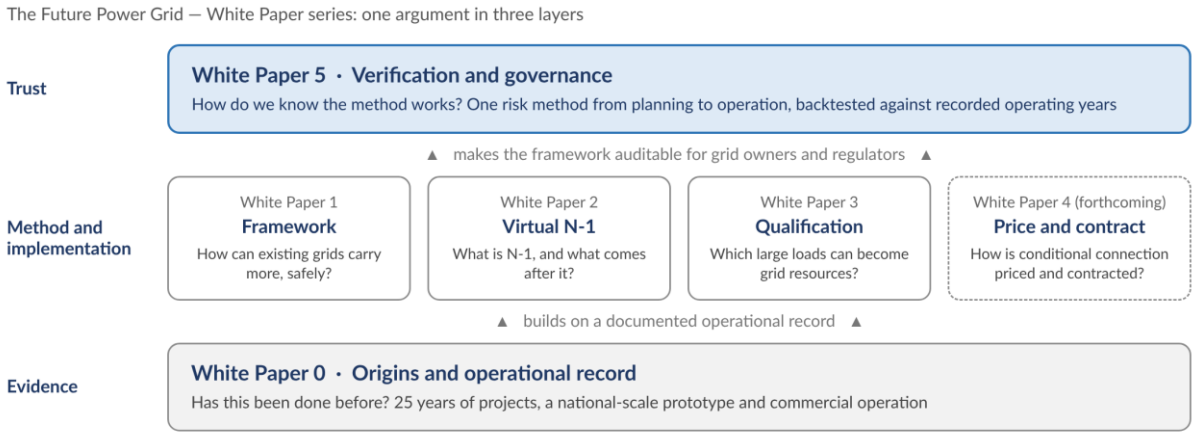


Figure 2 *The White Paper series in three layers: the operational record (White Paper 0) as evidence, White Papers 1–4 as method and implementation, and this paper as the verification and governance layer that makes the framework auditable. Conceptual illustration.*

2. One Main Task and One Main Goal

The customer of every grid owner is, in the end, the power system itself — from generation and transmission to distribution and end use. Grid owners are its stewards, responsible for it on behalf of society. Starting from the system rather than from a tool has a consequence: we first ask what the system must achieve, then what stands in the way, and only then which tools close the gap.

What the system must achieve is the same for every TSO and DSO in the world:

- **Main task:** keep generation and demand in continuous balance while operating within strict frequency and voltage limits.
- **Main goal:** deliver the required energy with the highest possible security of supply at the lowest overall cost.

The two are different kinds of statement. The task is deterministic: it is answered by load flow, contingency and dynamic stability analysis, which show whether the system withstands a defined event. The goal is probabilistic: it weighs security against cost, and that cannot be done without knowing how likely each event is and what it costs when it happens. White Paper 2 described the same distinction from the operating side: N-1 “is a risk criterion with no probability and no consequence”.

> ***An operator who documents only the task has not yet documented the goal. Traditional analysis documents the task; PRA documents the goal.***

Because the task and the goal are the same in every part of the value chain, and for every grid owner, the same assessment can in principle be applied everywhere. That observation is the foundation of the rest of this paper.

3. Why Now: Regulation Is Moving to Risk

The shift from deterministic to risk-based management is no longer a research ambition, as it was in European projects such as GARPUR (SINTEF, 2017). It is being written into regulation and national strategy, and the international energy institutions describe the same need.

Source	Requirement or finding	Timing
EU — ACER, CSAM Article 44	All TSOs jointly develop a common methodology for probabilistic risk assessment in operational security analysis, and propose it as an amendment to CSAM	31 Dec 2027
UK — The Grid We Need Now	Full probabilistic, risk-based grid operation and planning; probabilistic risk modelling of operation and risk metrics and thresholds by 2030	Operation 2035, planning 2036
UK — NESO and Ofgem, FRCR	Frequency-security policy set within the SQSS by weighing the cost of securing events against residual risk; FRCR 2025 approved after an independent review asked for clearer validation	In force since 2020; FRCR 2025 approved May 2026
IEA — Modernising Grids in the Age of Electricity	Operators keep margins below physical limits to cover risks they cannot fully observe; better information could narrow them safely; grid capacity must rise at least 30%	2035
United States — FERC, NERC	Risk focus mainly on resource adequacy and extreme weather (NERC, 2026); transmission planning criteria still largely deterministic	Ongoing

Table 2 *Regulatory and policy drivers for risk-based grid management*

The UK review states the problem plainly: deterministic standards involve "no weighing of probability – no consideration of how likely any particular failure is in reality, or the scale of its impacts". In Great Britain, constraint costs rose from £0.4 billion in 2019/20 to £1.7 billion in 2024/25, and the review notes that some curtailment occurred because headroom had to be kept against worst-case scenarios that were extremely unlikely at the time (UK Government, 2026).

Read together, the sources form one policy thread rather than separate initiatives. The IEA describes the need at global level: operators keep margins because they cannot fully observe and quantify risk (IEA, 2026b). ACER turns that need into an obligation for operation in Europe through CSAM. In Great Britain the Government review sets the national direction, and NESO and Ofgem already apply a probabilistic standard in practice: since 2020 the Frequency Risk and Control Report (FRCR) has set frequency-security policy within the Security and Quality of Supply Standard (SQSS) by weighing the cost of securing events against the residual risk (Ofgem, 2026b). Each step asks for the same two things: risk that is quantified, and evidence that can be checked.

Two observations from these sources shape this paper. First, the UK review notes that planning standards rely on assumptions about how the system will be operated, and therefore recommends moving operation first. CSAM likewise addresses operation. If operation becomes probabilistic by regulation while planning stays deterministic, the two will speak different languages for a decade or more. Second, the IEA finds that successful digital solutions often stall because each new deployment has to be justified again (IEA, 2026b). A method that is consistent across the value chain, and verifiable, removes much of that friction.

> ***If operation becomes probabilistic while planning remains deterministic, planning will make its promises in one language and operation will check them in another (authors' interpretation).***

4. Today's Analyses Show Half the Picture

Power system analyses are carried out in every part of a grid owner's value chain. They verify that the system withstands defined events. They do not say how likely those events are, or what the remaining risk costs.

Part of the value chain	Horizon	Typical decision
Investment	Decades	New lines, substations, connection capacity
Reinvestment	Years	Renewal or replacement of ageing assets
Maintenance planning	Weeks to months	Outage scheduling
Emergency preparedness	Days to months	Readiness for storms and extreme events
Operational planning	Hours to days	Transfer limits, reserves, switching plans
System operation	Now	Preventive and corrective actions

Table 3 *Where power system analyses are made in a grid owner's value chain*

Two weaknesses follow from today's practice:

- **Only the task is documented.** A deterministic N-1 check is binary. It answers "does the system withstand this?" but not "how likely is it, and what does it cost?". The main goal cannot be shown to be met.
- **No common thread.** Each part of the chain uses its own tools, criteria and assumptions. A promise made in a connection study cannot be checked against what happens in operation.

4.1 Deterministic Checks Hide Risk in Both Directions

The operational record shows that this is not a theoretical gap. In the 2008 study of new power supply to Troll A, load flow and voltage showed no problems in the grid as it was. The regularity analysis of the same grid showed that the BKK area was already at or beyond Statnett's proposed requirement of less than 15 system minutes in heavy load (White Paper 0, Section 4.2). The deterministic check and the probabilistic check gave different answers for the same grid.

The error also runs the other way. In substation design studies for National Grid (2008) and Landsnet (2009), quantified system risk showed that a simpler design could be justified where the surrounding network already provided redundancy (White Paper 0, Section 5). Deterministic rules hide real risk in some states and excess margin in others; quantified risk shows both.

4.2 "Operationally Justifiable" Without a Number

Under growing pressure from new connections, grid owners must decide whether a connection or an operating state is acceptable. In Norway the test is whether it is "operationally justifiable" (*driftsmessig forsvarlig*). The term is used in regulation but is not quantified. Without a number, the answer becomes a matter of judgement, and judgement under uncertainty tends to be conservative.

5. A Complete System Analysis in Four Steps

The established methodology for assessing new load already has the right structure. What is missing is a satisfactory answer to its third step.

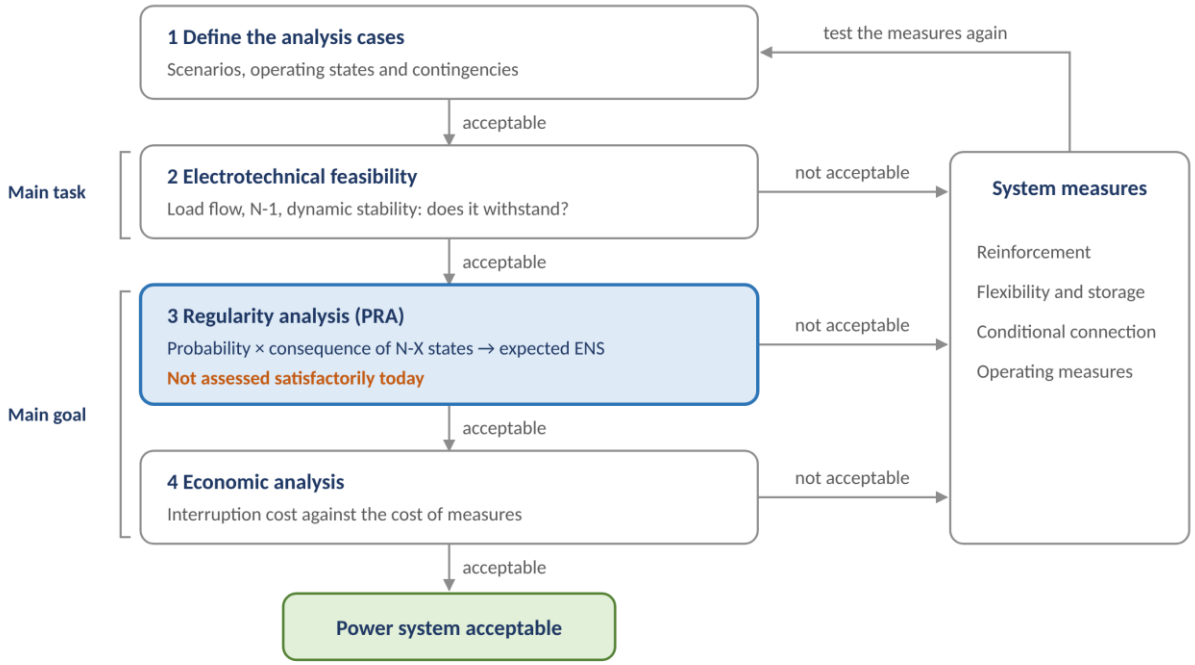


Figure 3 *A complete system analysis answers both the main task and the main goal. Step 2 documents the task; steps 3 and 4 document the goal. Step 3 is the step that is not assessed satisfactorily today.*

- **Step 1 — Define the analysis cases.** Scenarios, operating states and contingencies.
- **Step 2 — Electrotechnical feasibility.** Load flow, contingency and dynamic stability analysis. Does the system withstand the event?
- **Step 3 — Regularity analysis (PRA).** How likely is each event, including multiple outages, and what is the expected energy not supplied?
- **Step 4 — Economic analysis.** Expected interruption cost weighed against the cost of measures.

If any step is not acceptable, system measures are considered — reinforcement, flexibility and storage, conditional connection or operating measures — and tested again from step 1, until the power system is acceptable.

5.1 Making “Operationally Justifiable” a Quantity

With step 3 in place, risk is expressed as probability times consequence, aggregated to expected energy not supplied and its cost. Acceptance can then be set in two ways: relative to a reference (“no worse than today”), or by cost-benefit (“the measure costs less than the risk it removes”). Step 4 is where the main goal is made explicit, which is why the economic analysis belongs in the same consistent chain as steps 2 and 3.

Quantitative acceptance criteria are not new. The Troll A study referred to Statnett’s proposed requirement of less than 15 system minutes for the area (White Paper 0, Section 4.2). The UK review recommends developing risk metrics and thresholds by 2030 (UK Government, 2026). A framework for setting them can be proposed now.

5.2 Step 3 Has Been Accepted in Regulatory Practice Before

The Goliat electrification (2008–2010) shows that regularity analysis can form part of a regulatory decision. Eni’s power-system and regularity analyses were reviewed and verified by Statnett in dialogue with NVE, and formed

part of the documented basis for licensing (White Paper 0, Section 4.1). In the Troll A compressor project (2010–2011), the cost of energy not supplied was used as the common indicator across alternatives (White Paper 0, Section 4.5).

Step 3 does not require new planning standards first. PRA can be added to existing studies today, as a complement to deterministic analysis — not a replacement for it.

6. One Method Along the Whole Time Axis

Because the main task and the main goal are the same in every part of the value chain, steps 2 to 4 can be applied consistently from investment decisions to real-time operation (Figure 1).

6.1 What Consistency Means

Consistency means the same goal, the same model basis and the same acceptance criterion. It does not mean an identical calculation. Investment chooses lines and substations; operation chooses switching and redispatch. What changes along the time axis is the uncertainty of the inputs:

- **Years to decades:** scenarios for demand, generation and grid development.
- **Weeks to days:** weather forecasts, outage plans and market expectations.
- **Now:** measurements and the actual network state.

The UK review describes the same aim: risk-aware decisions "all the way through the stack, from contingency probability assessment at the planning layer to intraday re-dispatch at the operational layer" (UK Government, 2026). When this holds, what is promised in a planning or connection study can be confirmed in operation.

6.2 Operational Experience

One model serving several horizons has been demonstrated at distribution level. Dalane Nett's 2018 agreement covered operational use, maintenance prioritisation and reinvestment prioritisation on the same probabilistic model. In 2020, Lyse Elnett commissioned an interface between a real-time risk tool and its network model environment, with the stated objective of a uniform, holistic risk analysis across the value chain (White Paper 0, Sections 7.2–7.3). White Paper 1 calls this shared baseline the common operational reference frame.

> ***Only the input uncertainty changes along the time axis. The method, the model basis and the acceptance criterion stay the same — and that is what makes a planning promise checkable in operation.***

7. Verifiability: The Future We Predict, the Past We Verify

7.1 Trust Is the Barrier

Trust is the main barrier to risk-based management. The UK review illustrates why. In 2025 NESO, the system operator in Great Britain, proposed lowering the minimum inertia requirement on the basis of a probabilistic cost–risk assessment in its Frequency Risk and Control Report. The independent review published by Ofgem found the general approach to risk management sound, but the evidence "not presented sufficiently clearly to give stakeholders confidence in the level of risk to which the system is being exposed", and asked for clearer validation of the simulations and probability estimates (Ofgem, 2026a; UK Government, 2026). Ofgem declined to approve the change in December 2025. After NESO resubmitted with supplementary evidence, a phased reduction with extended trials and agreed success criteria was approved in May 2026 (Ofgem, 2026b). The IEA reports a similar pattern for AI in operation: explainability, auditability and accountability limit its use in the control room (IEA, 2026b).

The British case shows both halves of the argument. A probabilistic assessment can be accepted by a regulator; what delays acceptance is the lack of a clear, agreed way to validate it. The answer is to verify a PRA model against what actually happened, rather than to argue that it is right.

7.2 Replaying a Historical Operating Year

Once a PRA model quantifies expected energy not supplied minute by minute (Svendsen et al., 2012), the result can be aggregated over any period. A full historical operating year can then be taken from a grid owner's own database — the recorded topology, load, generation and weather — and run through the model again. Its predictions are compared with what was recorded.

7.3 Validation in Three Layers

Layer	Comparison	Statistical character
1. Disturbances	Expected number of disturbances from component failure rates against recorded disturbances, whether or not they led to lost load	Frequent events; a test with power month by month
2. Weather regimes	Weather-adjusted failure rates must predict when risk is high, not only the average; stormy months against calm months	Sharp test of the weather model
3. Energy not supplied	Expected against actual interruptions and their cost	Rare events dominated by a few large ones; needs several years

Table 4 *Three layers of validation for a PRA model (authors' proposal)*

Layers 1 and 2 validate the foundation of the model with good statistics. Layer 3 confirms the result over time.

7.4 Separating Model Error from Input Error

Replaying a year with the actual inputs removes forecast uncertainty, so the remaining deviation is the model's own. Replaying it with the forecast inputs available at the time shows how much is added by uncertainty in the inputs. Doing this at different horizons verifies each part of the value chain — and gives the planning-to-operation bridge its empirical basis.

7.5 What Has Already Been Demonstrated

The elements of this validation have each been demonstrated in practice, although not yet combined into a systematic backtest:

Element	Demonstration	Source
Replay of a recorded state	Møre blackout, 6 December 2014: the elevated risk state was visible for more than 19 hours; a study-mode replay of the stored state indicated that the loss of supply would most likely have been avoided. Presented to NVE as a hindcast	White Paper 0, Sections 7.1 and 7.3
Forecast against outcome	Storm Tor, 2016: pre-event forecasts of where outage risk was highest corresponded closely with the areas where more than 30,000 households lost supply	White Paper 0, Section 6.4
Out-of-sample test	Ten years of Icelandic weather data with a two-year blind test of a weather-dependent failure model	Svendsen et al. (2017)
Comparison with historical record	Landsnet's integrated model compared with the historical system-minutes average for 1998–2007	White Paper 0, Section 3.3

Table 5 *Elements of verification demonstrated in the operational record (White Paper 0)*

What has not yet been done is a systematic, year-long comparison of expected and recorded disturbances and energy not supplied for a whole grid. That is the natural next step.

7.6 Conditions for a Rigorous Test

- **Out of sample.** Failure rates are estimated on one period and tested on another; otherwise the model is checked against the data it was fitted to.
- **Common definitions.** An event in the model must mean the same as a recorded disturbance.
- **Dependent failures.** Storms that hit several components at once are where expected energy not supplied concentrates, and must be modelled explicitly.

Once PRA runs in real time, every new month becomes another out-of-sample test. The model builds a track record that a regulator can follow, instead of relying on a one-off argument. The Nordic countries are well placed to do this: decades of standardised fault and interruption statistics are available.

8. A Precedent from Banking

Banking supervisors faced the same problem thirty years ago. Since 1996, banks that use internal risk models for capital requirements have been required to backtest them against actual outcomes (Basel Committee on Banking Supervision, 1996). The UK review itself cites financial regulation as a sector that uses probabilistic risk alongside deterministic rules without compromising safety outcomes.

> ***"Conditioning supervisory standards on a clear framework, though limited and imperfect, is seen as preferable to a purely judgmental standard." — Basel Committee on Banking Supervision, 1996***

The Committee's reasoning applies directly to "operationally justifiable". The method is still in force in the current Basel Framework (MAR32, MAR99), and its structure translates closely to the grid:

	Banking (Basel market risk)	Power system (proposed)
Risk measure	One-day Value-at-Risk at 99%	Expected disturbances and expected energy not supplied
Compared with	Actual daily profit and loss, 250 trading days	Recorded disturbances and interruptions, month by month
Isolating model error	Hypothetical P&L: positions held constant, a pure test of the calculation	Replay with actual topology, load and weather
Including real-world effects	Actual P&L: includes intraday trading	Replay with forecast inputs
Verdict	Green 0–4, amber 5–9, red 10 or more exceptions	Zones from the same statistics, to be calibrated
Consequence	Amber: higher capital multiplier. Red: automatic increase, or model approval withdrawn	Amber: explain and recalibrate. Red: model not accepted as evidence

Table 6 *The Basel backtest translated to the power system*

The zones are set to balance two statistical errors: rejecting an accurate model, and accepting an inaccurate one. The Basel framework accepts that no single threshold keeps both errors low, so the amber zone asks the bank for more information before the supervisor acts (MAR99). A grid regulator could apply the same logic to a PRA model, with zone boundaries derived from the expected number of disturbances in the test period (authors' interpretation).

9. A Bridge Between Grid Companies

If every operator has the same task and the same goal, they can share a risk currency: expected energy not supplied and its cost, built on the same model basis and the same acceptance criterion.

- **TSO and DSO.** Switching, distributed generation and flexible demand in the distribution grid shift risk in the transmission grid, and the reverse. A common metric makes that transfer visible, hours or years ahead.
- **Across borders.** Coordinated security analysis and cross-zonal capacity are already shared tasks between TSOs. CSAM Article 44 gives them a common probabilistic basis, and ENTSO-E reports on the progress of that work every two years (ENTSO-E, 2025).
- **Towards the regulator.** One metric from planning to operation gives an audit trail: the risk accepted in an investment or connection decision can be followed into operation and checked.

Responsibilities and incentives still differ between TSOs and DSOs. A common metric does not remove those differences, but it shows when risk is moved across an interface instead of reduced.

10. From Promise to Contract: Closing the Loop

Conditional connection is proven practice. Goliat (2008–2010) accepted a large new load with load disconnection as a corrective measure under constrained conditions, and automatic disconnection of new compressor load became an explicit connection condition for Troll A in 2010 (White Paper 0, Section 4). White Papers 2 and 3 generalise this principle with quantified risk as the trigger. What has not been done is to close the loop from study to verification.

A large data centre connecting in phases shows the whole loop in one case:

- **Plan.** The connection study defines the conditions — battery storage, backup generation, grid-forming inverters and segmentation of the load — and quantifies the risk the system accepts with these conditions in place (steps 1–4).
- **Contract.** The conditions, and when they are activated, are written into the connection agreement (White Paper 4).
- **Operate.** Real-time PRA and DSA monitor the risk continuously and trigger the agreed measures when needed (White Papers 1–3).
- **Verify.** Backtesting compares the risk promised in the study with the risk observed in operation (this paper).
- **Learn.** The result of the backtest is fed back: failure rates are recalibrated, and acceptance criteria and triggers are adjusted for the next study. Planning, operation and verification then form one loop rather than a one-way chain (Figure 4).

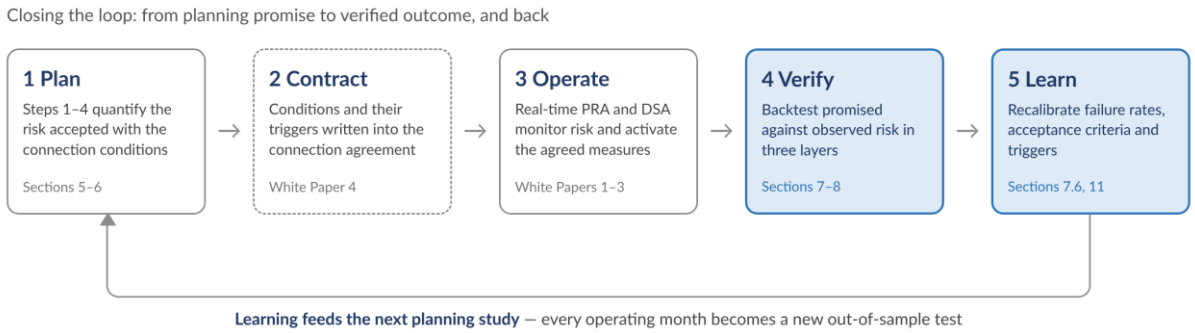


Figure 4 *Planning, contract, operation, verification and learning as one loop. Verification and learning are the contribution of this paper; the result of each backtest feeds the next planning study. Conceptual illustration.*

Verification is also what makes a conditional contract enforceable. If a connection agreement specifies a risk-threshold trigger, both parties need an agreed way to check afterwards that the trigger was set and used as intended. The backtesting method in Section 7 can provide that settlement basis, linking this paper directly to the price and contract terms of White Paper 4 (authors' interpretation).

11. Caveats and Qualifying Conditions

11.1 Acceptance Criteria Are Not Settled

The method makes “operationally justifiable” quantifiable, but it does not decide what level of risk is acceptable. That is a decision for grid owners, regulators and, ultimately, society. Who sets the criterion — the grid owner, the regulator, or the connection agreement — remains open.

11.2 Data Quality and Definitions

Backtesting is only as good as the recorded data. Fault statistics, interruption records and historical network states must be complete and defined consistently with the model. Where they are not, the first result of a backtest may be a data-quality finding rather than a model verdict.

11.3 Rare Events

Energy not supplied is dominated by a few large events. A single year cannot confirm or reject a model on that layer alone; layers 1 and 2 must carry the statistical weight in the short term.

11.4 Methodological Status

The three-layer validation, the translation of the Basel zones and the consistency principle along the time axis are proposals by the authors. The elements have been demonstrated separately in operation (White Paper 0), but a systematic backtest of a whole grid has not yet been carried out. The proposals should be tested on at least one grid before they are used in regulation.

12. Conclusions and Specific Recommendations

Every grid operator has the same main task and the same main goal. Traditional analysis documents the task; PRA documents the goal. Today the two are used in different parts of the value chain, with different tools and criteria, so a promise made in planning cannot be checked in operation. Regulation is now moving operation towards probabilistic risk, which will widen that gap unless planning follows.

This paper has proposed a governance and verification framework to close it: a four-step system analysis with PRA as the missing third step, one method along the whole time axis, and backtesting against recorded operating years to make the method verifiable. The elements have been demonstrated in operation. What remains is to make them consistent, systematic and accepted.

> ***The central claim: a risk method earns trust when its predictions can be checked against what happened. Applied consistently from planning to operation, and verified against history, quantified risk becomes something a regulator can audit — not only something an engineer can argue.***

The main challenge is no longer calculating risk. The challenge is creating a governance framework that allows quantified risk to be trusted, verified and regulated.

Specific Recommendations

- **For TSOs:** Add regularity analysis (step 3) to connection and planning studies now, using the same model basis as operation. Prepare the planning-side method so that it is consistent with the common operational PRA methodology developed under CSAM Article 44.
- **For DSOs:** Use one probabilistic model across operation, maintenance and reinvestment, as demonstrated at Dalane and Lyse. Make risk transfer at the TSO–DSO interface visible in the same metric.
- **For national regulatory authorities:** Define how “operationally justifiable” can be expressed as a risk acceptance criterion. Commission a pilot backtest of one historical operating year with one grid owner, with failure rates estimated out of sample.
- **For ACER and ENTSO-E:** Include verification — backtesting against recorded disturbances and interruptions — in the common probabilistic risk assessment methodology due by 31 December 2027.
- **For policymakers and international institutions:** Recognise verifiability as the condition for adoption. A risk method that can be checked against history removes the need to justify each deployment again.

12.1 Key Takeaway: Trust the Forecast You Can Check

Every day, weather services publish forecasts in probabilities — a 70% chance of rain — and every day those forecasts are checked against what actually happened. People plan their lives around a probabilistic statement because the forecaster’s record can be verified. Nobody asks the meteorologist to prove the model in advance. The track record does it.

Grid risk has rarely been treated this way. Deterministic rules are trusted because they are simple, not because their outcomes are checked. Probabilistic risk is doubted because it is new, even where it has a record.

The answer is to give grid risk the same discipline as the weather forecast: predict the future with one consistent method from planning to operation, and verify the past with the same method. It is difficult to make predictions, especially about the future. The past we can predict again — and check.

References

ACER. (2019, rectified). *Methodology for Coordinating Operational Security Analysis (CSAM), Annex I to ACER Decision 07/2019,* Articles 44 and 46. Agency for the Cooperation of Energy Regulators, Ljubljana.

Basel Committee on Banking Supervision. (1996, January). *Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements.* Bank for International Settlements, Basel.

Basel Committee on Banking Supervision. (2023). *Basel Framework, MAR32: Internal Models Approach – Backtesting and P&L Attribution Test Requirements.* Bank for International Settlements, Basel.

Basel Committee on Banking Supervision. (2022). *Basel Framework, MAR99: Guidance on Use of the Internal Models Approach.* Bank for International Settlements, Basel.

ENTSO-E. (2025). *Biennial Operational Probabilistic Coordinated Security Assessment and Risk Management Progress Report.* Brussels.

IEA. (2026a). *Electricity 2026.* Paris: International Energy Agency.

IEA. (2026b). *Modernising Grids in the Age of Electricity.* Paris: International Energy Agency.

NERC. (2026, January). *Long-Term Reliability Assessment.* North American Electric Reliability Corporation.

Ofgem. (2026a, March). *Frequency Risk and Control Report 2025: Independent Review.* Office of Gas and Electricity Markets, London.

Ofgem. (2026b, May). *Frequency Risk and Control Report 2025: Authority Decision.* Office of Gas and Electricity Markets, London.

SINTEF. (2017). *Generally Accepted Reliability Principle with Uncertainty Modelling and through Probabilistic Risk Assessment (GARPUR).* EU FP7 project 608540.

Svendsen, A. B. (2026). *Origins and Operational Record.* White Paper 0. The Future Power Grid. Draft.

Svendsen, A. B., et al. (2012). *Online Reliability Assessment of Power Systems.* PMAPS, Istanbul.

Svendsen, A. B., et al. (2017). *Modelling Weather Dependence in Online Reliability Assessment of Power Systems.* Journal of Risk and Reliability.

Svendsen, A. B., Nyiredy, R., & Milenkovic, M. (2026). *Unlocking Grid Capacity Through Risk-Based Operation.* White Paper 1. The Future Power Grid.

Svendsen, A. B., Nyiredy, R., & Milenkovic, M. (2026). *From N-1 to Virtual N-1: Extending the Risk-Based Framework for Grid Capacity.* White Paper 2. The Future Power Grid.

Svendsen, A. B., Nyiredy, R., & Milenkovic, M. (2026). *From Flexible Load to Grid Resource: A Qualification Framework for Large Flexible Energy Consumers.* White Paper 3. The Future Power Grid. Draft.

Svendsen, A. B., Nyiredy, R., & Milenkovic, M. (forthcoming). *Price and Contract for Conditional Connection.* White Paper 4. The Future Power Grid. Working title.

UK Government. (2026, September). *The Grid We Need Now: Independent Review of AI Deployment in the Electricity Networks.* London.