Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

DATA REQUIREMENTS FOR RELIABILITY-ORIENTED DIGITAL TWINS OF ENERGY SYSTEMS: A CASE STUDY ANALYSIS

Omar Mostafa[1] and Sanja Lazarova-Molnar[1,2]

[1] Institute AIFB, Karlsruhe Institute of Technology, Karlsruhe, GERMANY
[2] The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense, DENMARK

ABSTRACT

Ensuring reliability of energy systems is critical for maintaining a secure and adequate energy supply, especially as the integration of renewable energy increases systems' complexity and variability. Digital Twins offer a promising approach for data-driven reliability assessment and decision support in energy systems. Digital Twins provide decision support by dynamically modeling and analyzing system reliability using real-time data to create a digital replica of the physical counterpart. As modern energy systems generate vast amounts of data, it is essential to precisely define the data required for enabling Digital Twins for their reliability assessment. In this paper, we systematically investigate the data requirements for reliability-oriented Digital Twins for energy systems and propose a structured categorization of these requirements. To illustrate our findings, we present a case study demonstrating the link between data and model extraction for enhancing system reliability.

1 INTRODUCTION

Ensuring reliability of energy systems is important to maintain an adequate and secure energy supply, which has a direct impact on industrial productivity, economic stability and critical infrastructure. With the increasing share of Renewable Energy Sources (RES), the reliability of power systems is challenged by intermittent supply, unexpected disturbances and demand mismatches (Johnson et al. 2019; Denholm et al. 2020). While the impact of the probabilistic nature of RES on the adequacy of the power system has been well studied, more attention needs to be paid to the reliability of the renewable energy systems and power conversion systems (Niu et al. 2021). Unlike traditional power plants, which operate with predictable output, renewable energy generation is subject to variability due to weather, equipment degradation and external interactions. Failure of individual renewable components, such as inverters, battery storage or photovoltaic cells, can result in degraded performance or insufficient output (Sonawane et al. 2023). As managing renewable energy systems and ensuring the reliability of their electrical equipment becomes more complex, it is important to use advanced digital and data-driven techniques to assess the reliability of energy systems throughout their lifecycle (Li and He 2021).

The rapid evolution of digital technologies has paved the way for data-driven reliability assessments in energy systems. The integration of data-driven reliability assessment with Digital Twins (DTs) offers a transformative approach to improving reliability of future power systems (Song et al. 2023). DTs can enable real-time monitoring, predictive maintenance, and optimization capabilities (Li and He 2021). DTs can also address various challenges in smart energy systems, including digitization and socio-economic/environmental transitions and can enable remote monitoring, condition assessment, fault diagnosis, and optimization of renewable energy systems, transmission equipment and storage systems (Ardebili et al. 2021). Numerous research papers have demonstrated DTs' ability to improve reliability of various energy systems and revolutionize the energy sector (e.g., predicting failures and optimizing maintenance strategies) (Yu et al. 2022; Jafari et al. 2023). As the world faces the challenges of integrating

2025 Winter Simulation Conference (WSC) | 979-8-3315-8726-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/WSC68292.2025.11338875

renewable energy sources and optimizing energy consumption, the role of DTs in creating more sustainable and resilient energy systems has become increasingly important.

A key feature of DTs is their bidirectional connection to physical systems, with data flowing both from the physical system to the DT and vice versa. The underlying models of DTs rely on a continuous flow of high-quality data to accurately represent and simulate their physical counterparts (Glaessgen and Stargel 2012). The continuous exchange of information between the physical system and its DT allows for adaptive updates and dynamic feedback. This adaptability enables DTs to proactively respond to emerging issues and provide decision support to support a more stable, responsive, and sustainable energy infrastructures.

To enable the development and implementation of DTs in energy systems, it is essential to understand and address the data requirements that enable and facilitate their implementation and functionality. Even more importantly, data needs to be matched to the purpose of each specific DT (Lazarova-Molnar 2025). Our focus is on DTs for assessment of reliability of energy systems. Accuracy and effectiveness of DTs is directly related to the quantity, quality, and timeliness of the data they receive (Ebrahimi 2019).

To illustrate the data requirements for reliability-oriented DTs in energy systems, we present a case study of a small photovoltaic (PV) system. Based on literature insights, this case study identifies essential data required for reliability assessment (such as sensor data, fault records, and environmental factors) and demonstrates the extraction of reliability models from real system data. The study helps illustrate the theoretical concepts and data collection challenges in DT implementation.

In this case study, we aim to explore and define the essential data requirements for effective development and implementation of DTs for reliability assessment of energy systems. We begin by identifying essential system-level parameters (state and condition monitoring data) to model the reliability of the energy system (in this case a small photovoltaic system). Using these insights, we build a case study to demonstrate how a DT can use this data to reduce outages and improve fault analysis. Our case study is based on the Fault Tree (FT) reliability models that the DT automatically discovers from system data and then uses to assess system reliability. The simulated data includes FT basic events such as sensor data, fault records, and environmental factors that affect component performance.

In this paper, we begin with a literature review of the related work on energy system reliability using DTs in Section 2. Then we identify the data requirements for using DTs to maintain reliability of energy systems in Section 3. After that, we demonstrate the extraction of reliability models from energy system data with an illustrative case study in Section 4. Finally, we summarize our findings and discuss potential extensions of this work in Section 5.

2 BACKGROUND AND RELATED WORK

In the following, we provide a background on reliability assessment in energy systems, as well as an overview on the use of Digital Twins (DTs) in energy systems.

2.1 Reliability of Energy Systems

Reliability of energy systems has two fundamental facets: security and adequacy (Tuinema et al. 2020). Adequacy is the system's ability to meet demand under normal operating conditions, while security is the system's ability to withstand disturbances like outages or extreme weather events (Fulli et al. 2016). Each facet of reliability requires different categories of data to support the functionality of a reliability-oriented DT. For example, the data required to assess reliability of a component's availability (adequacy) differs from the data required to assess a component's remaining service life (security).

Traditional reliability assessments of energy systems are often characterized by limited automation and heavy dependence on manual methods. Reliability is typically evaluated using analytical methods using classical models (e.g., event trees, FTs, or reliability block diagrams) and simulation methods (e.g., Monte Carlo) (Hou et al. 2021). However, the increasing complexity and dynamic nature of modern energy systems, and rapidly evolving energy networks, incorporating renewable energy sources, smart grids and distributed energy resources, requires continuous and adaptive reliability assessment (Bera et al. 2020). To enable continuous and adaptive reliability assessment, automated reliability assessment methods are becoming increasingly relevant (Bertozzi et al., n.d.; Duchesne et al. 2020).

Fault Tree Analysis (FTA), for example, is a widely used reliability assessment technique that allows the identification of critical failure paths by analyzing the logical dependencies between component failures (Trivedi and Bobbio 2017). FTA consists of two main approaches: qualitative and quantitative. The qualitative approach focuses on analyzing the structure and components of the FT, while the quantitative approach focuses on calculating key metrics such as failure probabilities and system reliability using the FT (Trivedi and Bobbio 2017). However, FTA is typically based on expert-knowledge rather than observable data from the system (Niloofar and Lazarova-Molnar 2023).

Recent advances in FT extraction have focused on automated, data-driven methods. For example, Verkuil et al. (2022) used the C4.5 decision tree and LIFT algorithm (Learning FTs from Observational Data) to generate explainable FTs from sensor data (Nauta et al. 2018). Grimmeisen et al. (2022) introduced a case study on model-to-model transformation to derive FTs from DTs, integrating them with Markov chains for continuous reliability assessment. Niloofar and Lazarova-Molnar (2023) introduced their DDFTA algorithm, using a naive Bayes classifier to predict failures from time series data. These approaches improve the adaptability and accuracy of reliability assessment for complex power and industrial systems.

2.2 Digital Twins for Reliability of Energy Systems

DTs are high-fidelity digital replicas of physical systems that continuously collect and analyze data for informed decision making. The concept originated at NASA in the 1960s as a “living model” for the Apollo program (Allen 2021) and was later introduced to industry by Michael Grieves in 2002 (Grieves and Vickers 2017). In 2012, NASA defined DTs as integrated, multiphysics, multiscale, and probabilistic simulations of a system that use physical models, sensor updates, and operational history to mirror its corresponding physical twin (Glaessgen and Stargel 2012). Since then, the concept of DT has evolved to use advanced technologies such as smart sensors, smart devices, cloud platforms, artificial intelligence (AI), and the Internet of Energy (IoE) in electric grids (Sifat et al. 2023).

In energy systems, DTs serve as virtual entities that replicate the properties, behaviors, and interactions of physical energy assets such as power plants, grids, and substations (Song et al. 2023). They enable real-time monitoring, performance optimization, and predictive maintenance by creating a continuous feedback loop between the physical and virtual worlds (Palensky et al. 2022; Pan et al. 2020). DTs are transforming energy systems by enhancing efficiency, reliability, and sustainability across various sectors, including power generation, transmission, distribution, energy storage, industrial management, and smart cities (Mchirgui et al. 2024). DTs are increasingly used in various facets of energy systems to improve cybersecurity, efficiency, sustainability, and reliability (Jafari et al. 2023; Cali et al. 2023).

Other studies have demonstrated the capabilities of DTs in virtual simulation, condition monitoring, performance optimization, and fault diagnosis for renewable energy systems (Li and He 2021). For example, De Kooning et al. (2021) provide a comprehensive review of modeling techniques for wind turbine components in the context of DTs for wind energy conversion systems. Similarly, Augustyn et al. (2021) presents a probabilistic framework for updating the structural reliability of offshore wind turbine substructures with DTs. However, these studies focus primarily on component modeling and structural reliability, respectively, without considering the overall system reliability.

Despite significant advances in DT research, there remains a research gap in the existing literature regarding the development and implementation of DTs for energy system reliability. This gap highlights the need for more focused research efforts to bridge the gap between DT technology with advanced reliability assessment methods for energy systems. While the potential benefits of DTs for improving energy reliability are clear, the data required to automatically create accurate reliability models (i.e., digital replicas) are not well defined. This limitation underscores the need for further research into the data requirements for DTs in power systems, through case study analysis, which can provide practical insights into implementation challenges and best practices.

3 DATA REQUIREMENTS FOR RELIABILITY-ORIENTED DIGITAL TWINS OF ENERGY SYSTEMS

Data generated by energy systems can be used as input to DTs to simulate system behavior and enable informed decisions to improve reliability. However, the effectiveness of DTs in improving energy system reliability is fundamentally dependent on the quantity, quality and timeliness of the data provided. Therefore, to develop and implement reliability-oriented DTs for energy systems, it is essential to identify and categorize the necessary data sources. This section defines the data required for DT implementation and their potential sources.

3.1 Reliability Models for Energy Systems

To determine the data requirements for DTs aimed at enhancing the reliability of energy systems, it is first necessary to understand the underlying reliability models that are used to evaluate them. Reliability assessment is typically performed using either analytical models or simulation models. Tuinema et al. (2020) outline three main categories of reliability modeling relevant to energy systems: components, small systems, and large systems. While small systems allow for component-level modeling, large systems require aggregation approaches due to their complexity and scale. Below is an overview of reliability models for energy systems.

3.1.1 Reliability Models of Components in Energy Systems

Component reliability models form the basis of reliability analysis for energy systems. Typical approaches to component reliability modeling include the use of probability distributions such as the exponential, Weibull, or bathtub curve (Trivedi and Bobbio 2017). Each distribution is represented by a Probability Density Function (PDFs) done through parameters such as failure rate (λ) and repair rate (μ). The exponential distribution is widely used to describe random failures characterized by a constant failure rate (steady-state operation) such as an electronic component like circuit breakers and relays. The PDF of an exponential distribution is $f(t) = \lambda e^{-\lambda t}$, where λ is the constant failure rate and t is time. The Weibull distribution, on the other hand, can describe components with non-constant failure rates (age-dependent operation) such as a mechanical component like wind turbine gearboxes and bearings. The PDF of a Weibull distribution is expressed as

$$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^{\beta}},$$

where β is a shape parameter ($\beta \geq 0$), and η is a scale parameter ($\eta \geq 0$). For a shape parameter $\beta < 1$, $\beta = 1$, and $\beta > 1$, the failure rate is decreasing, constant, and increasing over time, respectively.

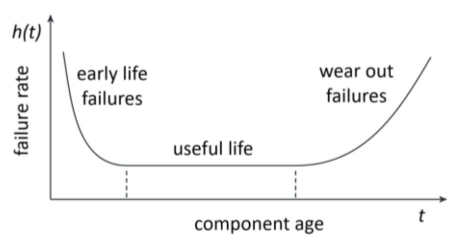

Figure 1: The Bathtub Curve.

The Bathtub curve (Figure 1) describes component failure rates over the entire life cycle: an initial high rate from early defects, a constant rate from random failures, and a rising rate in the wear-out phase due to aging and degradation.

Another approach to component reliability modeling is the two-state Markov model (Tuinema et al. 2020), which represents components in two states, either available (up) state or unavailable (down) state, suitable for repairable systems. The probabilities of these states are represented by:

$$P_{S_{up}} + P_{S_{down}} = 1, P_{S_{up}} = \frac{\mu}{\lambda + \mu} = A, P_{S_{down}} = \frac{\lambda}{\lambda + \mu} = U,$$

where $P_{S_{up}}$ the probability of the up state which equals the availability A of the component, and $P_{S_{down}}$ the probability of the down state which equals the unavailability U of the component. Therefore, to model the reliability of components, data on failure times and repair times is required.

3.1.2 Reliability Models of Small Energy Systems

Small energy systems consist of several components. Common reliability models for small energy systems are reliability networks, FTs, event trees and Markov models. For example, Fault Tree Analysis (FTA) systematically identifies potential failure modes and their logical links to system-level failures, offering qualitative insight into critical failure paths and quantitative metrics such as system reliability and failure probabilities (Tuinema et al. 2020). In FTA, a top event is the undesired system failure being analyzed, intermediate events are failures caused by other events, and basic events are the simplest occurrences that represent component-level faults. System failure probabilities are derived from basic events probabilities using Boolean logic (Niloofar and Lazarova-Molnar 2023). Markov models are also effective for reliability modeling of small energy systems, especially for capturing various states of the same components (Tuinema et al. 2020). Thus, modeling reliability of small systems requires data on component failure probabilities, system architecture/topology, and interdependencies between components.

3.1.3 Reliability Models of Large Energy Systems

For larger energy systems, modeling each component state would result in an extremely complicated model. Therefore, there are reliability methods specific to larger systems, such as state enumeration and Monte Carlo simulations. State enumeration considers system states defined by different combinations of component states to determine failure probabilities and impacts. When analytical enumeration is impractical, Monte Carlo simulation can be used. Monte Carlo simulation estimates system reliability by simulating random failure scenarios to evaluate their impact on system reliability. Both methods are used to calculate probabilistic reliability indicators such as Loss of Load Probability (LOLP) or Expected Energy Not Supplied (EENS). These methods typically require data on component failure rates, operational loads, and generation profiles to accurately model system reliability (Tuinema et al. 2020).

3.2 Categorizing Data for Enabling Reliability-Oriented Digital Twins in Energy Systems

Reliability-oriented DTs in energy systems require a variety of data to support the modeling techniques introduced in Section 3.1. Kasper et al. (2022) emphasize that effective DT platforms must meet the specific data and integration needs of industrial energy systems. The data needed for automatic reliability assessment of an energy system can be grouped into state data and condition monitoring data. Both are represented as time series and are critical for learning reliability models of energy systems.

State data captures discrete states of system components over time, such as operating states and fault events. Typically in the form of fault records of system components, this data enables automatic learning of systems' reliability models, such as a FTs (Niloofar and Lazarova-Molnar 2023; Dai et al. 2022). Condition monitoring data, on the other hand, includes continuous sensor readings, such as temperature, voltage, vibration, and pressure, providing insights into component health. This data can be used to automatically detect degradation patterns and estimate failure rates through probability distribution fitting or machine learning (e.g., Weibull or exponential models) (Friederich et al. 2021; Li and He 2021).

Based on the reviewed literature, we categorize the data sources essential for reliability-oriented DTs for energy systems into four application-specific categories: component-level, system-level, environmental data, and expert knowledge. **Component-level data** include component's operating parameters and performance metrics used to estimate component health and failure rates. **System-level data** capture interactions between components and system performance metrics to identify potential failure modes for small system as noted in Section 3.1. Examples include system topology, unit capacity, and load. **Environmental data** capture external factors affecting system reliability, such as weather conditions and grid stability information. This kind of data allows DT models to respond to changing external conditions, improving overall reliability. Finally, **expert knowledge** supports validation of DTs' underlying reliability models with respect to behaviors of corresponding real-world systems. Figure 2 illustrates these four data categories hierarchically according to the scope of integration and relevance to reliability-oriented DTs.

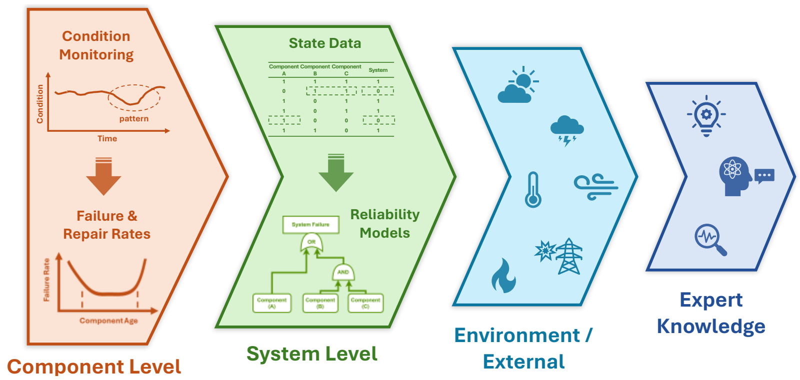

Figure 2: Levels of data sources essential for reliability-oriented DT in energy systems specific to the application.

3.3 Linking Reliability Models to Data for Reliability-Oriented Digital Twins of Energy Systems

This section links the reliability models discussed in Section 3.1 with the corresponding data requirements outlined in Section 3.2. Table 1 outlines the specific data needed for reliability-oriented DT in energy systems, based on different reliability models, allowing for automated model extraction and adaptation.

4 ILLUSTRATIVE CASE STUDY

In the following, we present a case study to illustrate the data requirements for extraction of reliability models from energy system data to enable Fault Trees (FTs) as underlying models for reliability-oriented DTs for energy systems. The goal is to identify the data needed to accurately reconstruct the original reliability model from a state log using a data-driven method. For FT model extraction, we use the Data-Driven Fault Tree Analysis (DDFTA) method, as introduced by Niloofar and Lazarova-Molnar (2023).

4.1 Case Study Model

Our case study examines a small solar power system consisting of a Photovoltaic (PV) module, a diode, a Miniature Circuit Breaker (MCB), a fuse, and an electrical load. The system converts solar energy to electricity, making it suitable for studying the impact of solar variability on reliability. The diode provides unidirectional current flow, while the fuse protects against overcurrent from the PV module. We assume a direct current (DC) load, such as a battery storage system. Figure 3a illustrates the case study system layout.

The reliability model is a FT, which typically consists of three levels: basic events (BE), intermediate events (IE) and a top event (TE) (Trivedi and Bobbio 2017). BEs include electrical failures, material degradation, high temperatures, environmental factors (e.g., shading) and other technical faults. IEs are derived from BEs and reduction in the power output or performance of components. TE represents the reduction or failure of the system to generate energy and serves as the ultimate indicator of reliability (Sonawane et al. 2023). Figure 3b shows the FT model for the system, which is also the ground truth model.

Table 1: Linking data requirements with different reliability models in energy systems.

Reliability Model	Data Required	Model Extraction	Key Performance Indicators (KPIs)	Energy System Examples	References
Reliability Distributions	Maintenance logs (failure and repair times/duration)	Distribution Fitting (exponential, Weibull, bathtub, etc.),	Failure rate, Mean-Time-To-Failure (MTTF), Mean-Time-To-Repair (MTTR)	Components: Battery storage, transformer, turbine gearbox	(Trivedi and Bobbio 2017)
Reliability Block Diagram (RBD) & Reliability Networks	System architecture/topology, interdependencies, component failure probabilities	Topological analysis of system connections from minimal cut sets	Overall system reliability and failure probability, risk indices, sensitivity indices	Small-scale grids, distribution networks, microgrids	(Tuinema et al. 2020)
Markov Models/ Two-State Markov Model	Component state data (up/down), failure/repair times, sensor and operational time-series data	Hybrid physics-guided neural network modeling of state transitions, variational inference training	Availability, MTTF, MTTR, transition probabilities	Components: Battery storage, inverter, transformer	(Liu et al. 2022)
Fault & Event Tree Models	Basic event failure probabilities, system architecture/topology, interdependencies	Topological analysis of system connections from minimal cut sets	Overall system reliability and failure probability, risk indices, sensitivity indices	Small-scale grids, distribution networks, microgrids	(Lazarova-Molnar et al. 2020; Niloofar and Lazarova-Molnar 2023a; Sonawane et al. 2023)
Large System Models (State enumeration and Monte Carlo simulations)	Time-series load and generation data, component failure data, outage records weather/environment	Data-driven probabilistic methods	Voltage stability indices, overload probabilities, reliability indices (e.g., LOLP, EENS)	Large-scale networks, smart grids, transmission networks	(Tuinema et al. 2020)

We use the ground-truth FT model of the PV system (Figure 3a) and the failure probabilities in Table 2 to generate synthetic data for faults and failures, including event logs, sensor readings, and environmental conditions that reflect the real-world behavior. Niloofar and Lazarova-Molnar (2023) introduced a Data-Driven FT Analysis (DDFTA) method, which automatically learns FTs from time-series fault data. Using this method, we automatically construct the FT from the synthetic data. To validate the extracted FT model, we compare it to the ground-truth FT model either through its Boolean equivalent or truth table. The Boolean expression represents the FT structure through logical relationships between components and system failure (TE), while the truth table lists all possible component states and their impact on the TE. This validation ensures accuracy of the data-driven reliability model extraction.

From the original FT model of the PV system (Figure 3b), the Boolean expression can be derived using the plus sign (+) to represent logical OR gates and the dot (·) to represent logical AND gates between BEs. The resulting Boolean expression of the ground-truth FT model is given in equation (1):

$$TE = BE_1 + BE_2 + (BE_3 \cdot BE_4) + BE_5 + \cdots + BE_{17} + BE_{18} \tag{1}$$

4.2 Illustration of Data Requirements

To extract the system reliability model from data, this data must be first identified. The original FT model of the system includes several component faults and failure modes and their relationships with the system failure. To derive these relationships, we need to analyze the possible faults/failures (internal or external) and study their impact on the system performance. Therefore, the data needed for such a study include time-series fault logs and sensor measurements related to system performance.

Fault logs are historical records of discrete events and alarms that indicate specific component faults or failures detected by monitoring systems. However, fault logs often overlook gradual performance degradation, such as soiling, partial shading, or hot spots in PV cells, that impact overall system reliability without triggering alarms. Continuous data from sensors measuring voltage (V), current (I), and temperature can help detect component performance degradations. For example, a localized temperature spike in a PV cell compared to other PV cells may indicate a developing hot spot in the PV module. Integrating sensor data with fault logs provides a complete view of system reliability, enabling more accurate and proactive reliability modeling and analysis.

The data required in our case study include fault logs, sensor readings from the system (e.g., voltage, current, temperature), and environmental conditions (e.g., solar irradiance). These data can be used to detect fault/failure events and their relationship to system performance (i.e., to extract reliability models such as FT) collected from sources such as Supervisory Control and Data Acquisition (SCADA) systems or Industrial Internet of Things (IIoT) devices. However, collecting accurate and complete data can be challenging due to gaps in fault logs and limited access to SCADA and IIoT devices. To generate the synthetic data, we used literature-sourced failure probabilities for each of the BEs in the original FT model.

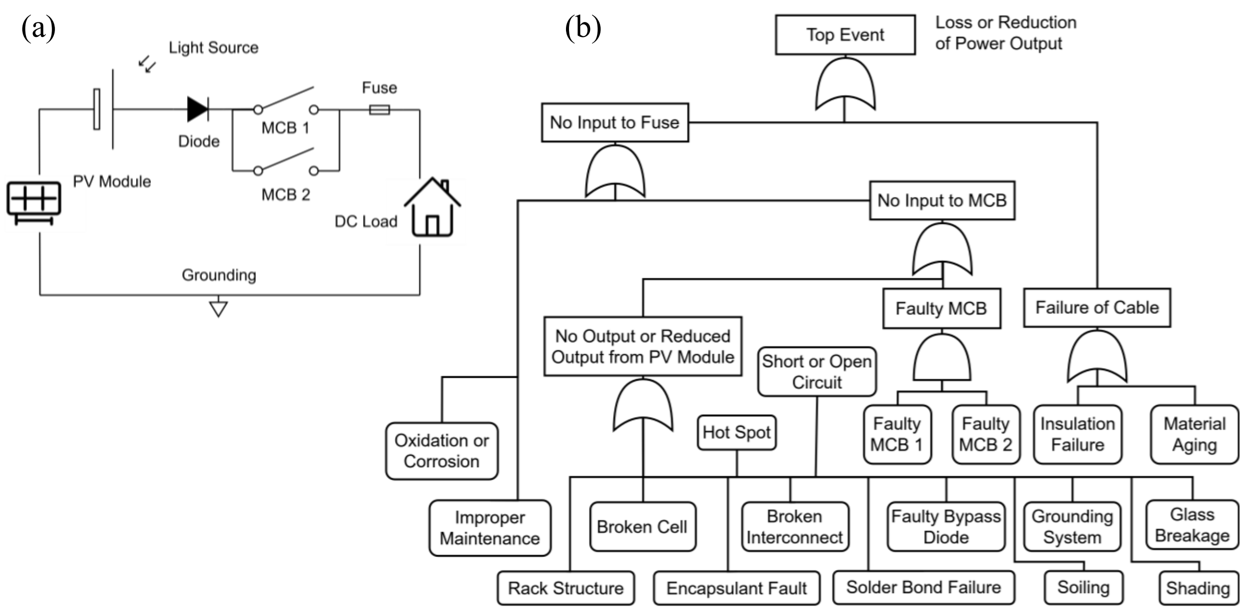

Figure 3: (a) PV system layout for the case study. (b) True FT model of the case study system.

4.3 Model Extraction from Data

For our case study, we used data on failure probabilities from the existing literature on PV system components and subcomponents such as cables, racks, and grounding (Sonawane et al. 2023; Colli 2015; Golnas 2013). For simplicity, all faults and failures are assumed to have constant failure rates following the exponential distribution. This assumption does not limit the model extraction approach, which supports arbitrary probability distributions. Table 2 lists fault/failure occurrence probabilities, data requirements, and data sources for reliability model extraction for all the BEs of the original FT model (Figure 3b). We generated synthetic fault logs using failure probabilities of BEs and Monte Carlo simulations to create a synthetic fault log for the PV system, mimicking real-world data. The synthetic data are simplified/reduced into state data serve as an input for the FT extraction algorithm. The state data represents the truth table of the original FT model of the PV system, which we use later to validate the extracted FT model result. Table 3 shows an example of the generated fault log and its equivalent state data shown in Table 4.

Table 2: Faults and Failures in the Case Study PV System: Associated Occurrence Probabilities, Data Requirements, Data Sources, and Data Categories for Reliability Model Extraction.

System Element	Fault / Failure	Basic Event	Occurrence Probability (from literature)	Data Required for Reliability Model Extraction	Possible Data Sources (from actual systems)	Data Source Category
Fuse	Oxidation or Corrosion	(BE_1)	0.0001	Electrical resistance or temperature	Ohmmeter, thermal sensor	Component-level
	Improper Maintenance	(BE_2)	0.0002	Maintenance records	Maintenance Logbooks	Expert Knowledge
MCB	Faulty MCB	(BE_3, BE_4)	0.0008	Voltage or current	I-V sensor over MCB	Component-level
PV Module	Broken Interconnect	(BE_5)	0.0846	Thermal images or	Infrared camera, I-V curve tracer	Component-level
	Grounding System	(BE_6)	0.0490	Ground insulation resistance or leakage current	Ground-fault detector	System-level
	Glass Breakage	(BE_7)	0.0003	Visual inspection	Camera	Component-level
	Soiling	(BE_8)	0.0013	Irradiance and PV output	Pyranometer, I-V curve tracer	Environment / External
	Shading	(BE_9)	0.0088	Irradiance	Pyranometer	Environment / External
	Broken Cell	(BE_{10})	0.1115	Thermal imaging and cell power output	Infrared camera, cell I-V tracer	Component-level
	Solder Bond Failure	(BE_{11})	0.1487	Cell power output	Cell I-V tracer	Component-level
	Hot Spot	(BE_{12})	0.0101	Thermal imaging	Infrared camera	Component-level
	Faulty Bypass Diode	(BE_{13})	0.0021	Voltage or current	I-V sensor over diode	Component-level
	Short/Open Circuit	(BE_{14})	0.0052	Voltage or current	I-V sensor	System-level
	Rack Structure	(BE_{15})	0.0729	Tilt angle, orientation, or vibration	Gyroscope or camera	System-level
	Encapsulant Fault	(BE_{16})	0.0570	Visual inspection and PV output	High-resolution camera, I-V curve tracer	Component-level
Cable	Insulation Failure	(BE_{17})	0.0001	Electrical insulation resistance	Ohmmeter	Component-level
	Material Aging	(BE_{18})	0.0002	Cumulative environmental exposure	Weather station data (UV index, temperature)	Environment / External

Table 3: Exemplary fault log dataset for a PV system generated using failure probabilities.

Timestamp	Component	Description	Status	Severity
2025-02-12 10:53:24	PV_Grounding_System_Fault	Grounding System Fault detected.	Active	High
2025-02-12 17:32:11	PV_Grounding_System_Fault	Grounding System Fault detected.	Cleared	High
2025-02-17 11:20:54	PV_Panel_Fault	PV Hot Spot detected.	Cleared	Medium
2025-02-18 13:23:51	PV_Panel_Fault	PV Solder Bond Failure detected.	Active	High
2025-02-18 08:55:02	Structure_Fault	Structural misalignment.	Active	Low

Table 4: State dataset based on the synthetic fault logs from Table 3.

Timestamp	Grounding System	Hot Spot	Solder Bond Failure	Rack Structure	Loss or Reduction in Power Output
1	1	0	0	0	1
2	0	0	0	0	0
3	0	1	0	0	1
4	0	0	0	0	0
5	0	0	1	0	1
6	0	0	0	1	1

To extract FT model in our case study, we have implemented an algorithm using Python that automatically identifies Minimal Cut Sets (MCSs), smallest combinations of BEs (component failures) leading to the TE, from the truth table (state data). The algorithm iterates over the state data to find MCSs and then derives a Boolean expression that captures the logical structure of the FT model. For validation, we compare the extracted Boolean expression to the original FT model (Figure 3b), represented by Equation (1). Algorithm 1 outlines this process. The code used is publicly available (Mostafa 2025). Algorithm 1 shows the extracted FT model from data using the algorithm. Validating the Boolean expression in Figure 4 with equation (1), we find that both the original model and the extracted FT model match.

Algorithm 1: Extracting FT Boolean Expression from Truth Table.

Input: $\mathcal{T} \in \{0,1\}^{m\times(n+1)}$: Truth table with m rows $(R_1, R_2, \dots, R_m)$ timestamps, and $n+1$ columns $(BE_1, BE_2, \dots, BE_n, TE)$.
Output: $\Phi(BE_1, BE_2, \dots, BE_n)$: Extracted FT Boolean expression.

Load $\mathcal{T}$	➢ Load the truth table as input dataset $\mathcal{T}$.
Step 1: Identify Cut Sets	
Initialize $CS \leftarrow \emptyset$	➢ Create dataset CS to store potential cut sets.
foreach $i \in \{1, 2, \dots, m\}$ **do**	➢ Iterate over rows in the $\mathcal{T}$ to identify cut sets.
if $\mathcal{T}[i, TE] = 1$ **then**	
$C_i \leftarrow \{BE_i \mid R_i[BE_i] = 1\}$	➢ Check for set of BEs that lead to $TE = 1$.
$CS \leftarrow CS \cup \{C_i\}$	➢ Store identified set of BEs as a cut set in CS.
Step 2: Minimalize Cut Sets	
Initialize $MCS \leftarrow \emptyset$	➢ Creating a dataset MCS to store minimal cut sets.
foreach $C_i \in sort(CS, by\ \lvert C\rvert \uparrow)$ **do**	➢ Iterate over all cut sets in CS to scan for minimal cut sets.
if $\nexists\ M \in MCS\ such\ that\ M \subseteq C_i$ **then**	➢ Identify minimal cut sets by eliminating supersets from CS.
$MCS \leftarrow MCS \cup \{C_i\}$	➢ Store identified minimal cut sets into MCS.
Step 3: Construct Boolean Expression	
foreach $M_i \in MCS$ **do**	➢ Construct the Boolean expression using logical OR $(+)$ between minimal cut sets and logical AND $(\cdot)$ between BEs.
$\Phi = \sum_{Mi \in MCS} \prod_{BE \in M_i} BE$	

```
Identifying Cut Sets (first 4 shown):   Minimal Cut Sets:
    Cut Set 1: ['BE18']                  ['BE18'], ['BE17'], ['BE16'], ['BE15'], ['BE14'],
    Cut Set 2: ['BE17']                  ['BE13'], ['BE12'], ['BE11'], ['BE10'], ['BE9'],
    Cut Set 3: ['BE17', 'BE18']          ['BE8'], ['BE7'], ['BE6'], ['BE5'], ['BE2'], ['BE1'],
    Cut Set 4: ['BE18', 'BE4']           ['BE3', 'BE4']

Extracted Fault Tree Boolean Expression:  TE =  BE18 + BE17 + BE16 + BE15 + BE14 + BE13 + BE12 +
BE11 + BE10 + BE9 + BE8 + BE7 + BE6 + BE5 + BE2 + BE1 + BE3·BE4

Constructed Truth Table (sample):
R   BE1  BE2  BE3  BE4  BE5  BE6  BE7   ...   BE13  BE14  BE15  BE16  BE17  BE18  TE
0   0    0    0    0    0    0    0     ...   0     0     0     0     0     0     0
1   0    0    0    0    0    0    0     ...   0     0     0     0     0     1     1
2   0    0    0    0    0    0    0     ...   0     0     0     0     1     0     1
3   0    0    0    0    0    0    0     ...   0     0     0     0     1     1     1
4   0    0    0    1    0    0    0     ...   0     0     0     0     0     0     0
[5 rows x 19 columns]
Truth table generated and saved from Boolean expression.
[Validation Successful]: Truth Tables Match!
The constructed fault tree produces an identical truth table to the original.
```

Figure 4: Extracted FT model using Algorithm 1.

5 SUMMARY AND OUTLOOK

We investigated the data requirements for the development and implementation of Digital Twins for the reliability assessment of energy systems. Specifically, we identified the distinct data categories, component-level data, system-level data, environmental data and expert knowledge, required to support effective Digital Twin applications. Using a photovoltaic energy system as an illustrative case study, we demonstrated that the combination of fault logs with continuous sensor data enables the automated extraction of data-driven reliability models, in this case Fault Trees. The fault logs of the photovoltaic system, which were synthetically generated using literature-sourced component failure probabilities, were transformed into structured state data that enabled Fault Tree model extraction. The extracted reliability models from Digital Twins can be used for reliability assessment of energy systems.

Our research shows that while the accurate extraction of reliability models, as enabler for reliability-oriented Digital Twins, is feasible, it presents challenges, particularly in system complexity and collecting time series, high quality, continuous data about the system. Further research is needed to address model validation and practical decision support to develop a standard framework for implementing reliability-oriented Digital Twins. Future studies could extend on our data-driven methodology by exploring alternative energy system applications and reliability models beyond Fault Trees, such as Reliability Block Diagrams and Markov models. Developing further methods for data-driven reliability model extraction would support the implementation of Reliability-Oriented Digital Twins in complex energy systems.

REFERENCES

Allen, B. D. 2021. "Digital Twins and Living Models at NASA." Document No. 20210023699, Langley Research Center (LaRC).

Ardebili, A. A., A. Longo, and A. Ficarella. 2021. "Digital Twin (DT) in Smart Energy Systems - Systematic Literature Review of DT as a Growing Solution for Energy Internet of the Things (EIoT)." *E3S Web of Conferences* 312:09002.

Augustyn, D., M. D. Ulriksen, and J. D. Sørensen. 2021. "Reliability Updating of Offshore Wind Substructures by Use of Digital Twin Information." *Energies* 14(18):5859.

Bera, A., A. Chowdhury, J. Mitra, S. Almasabi, and M. Benidris. 2020. "Data-Driven Assessment of Power System Reliability in Presence of Renewable Energy." In *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, August 18th-21st. Liege, Belgium.

Bertozzi, O., H. R. Chamorro, E. O. Gomez-Diaz, M. S. Chong, and S. Ahmed. n.d. "Application of Data-Driven Methods in Power Systems Analysis and Control." *IET Energy Systems Integration* 6(3):197–212.

Cali, U., B. D. Dimd, P. Hajialigol, A. Moazami, S. N. G. Gourisetti, G. Lobaccaro, and M. Aghaei. 2023. "Digital Twins: Shaping the Future of Energy Systems and Smart Cities through Cybersecurity, Efficiency, and Sustainability." In *2023 International Conference on Future Energy Solutions (FES)*, June 12th-14th. Vaasa, Finland.

Colli, A. 2015. "Failure Mode and Effect Analysis for Photovoltaic Systems." *Renewable and Sustainable Energy Reviews* 50:804–9.

Dai, P., L. Yang, Y. Zeng, M. Niu, C. Zhu, Z. Hu, and Y. Zhao. 2022. "Data-Driven Reliability Evaluation of the Integrated Energy System Considering Optimal Service Restoration." *Frontiers in Energy Research* 10.

De Kooning, J. D. M., K. Stockman, J. De Maeyer, A. Jarquin-Laguna, and L. Vandevelde. 2021. "Digital Twins for Wind Energy Conversion Systems: A Literature Review of Potential Modelling Techniques Focused on Model Fidelity and Computational Load." *Processes* 9(12): 2224.

Denholm, P., T. Mai, R. W. Kenyon, B. Kroposki, and M. O'Malley. 2020. "Inertia and the Power Grid: A Guide Without the Spin." NREL/TP-6A20-73856. National Renewable Energy Lab. (NREL), Golden, CO (United States).

Duchesne, L., E. Karangelos, and L. Wehenkel. 2020. "Recent Developments in Machine Learning for Energy Systems Reliability Management." *Proceedings of the IEEE* 108(9):1656–76.

Ebrahimi, A. 2019. "Challenges of Developing a Digital Twin Model of Renewable Energy Generators." In *2019 IEEE 28th International Symposium on Industrial Electronics (ISIE)*, June 12th-14th, Vancouver, BC, Canada:1059–66.

Friederich, J., S. C. Jepsen, S. Lazarova-Molnar, and T. Worm. 2021. "Requirements for Data-Driven Reliability Modeling and Simulation of Smart Manufacturing Systems." In *2021 Winter Simulation Conference (WSC)*, 1–12 https://doi.org/10.1109/WSC52266.2021.9715410.

Fulli, I. G., F. Profumo, and E. Bompard. 2016. *Electricity Security: Models and Methods for Supporting the Policy Decision Making in the European Union.* Ph.D. thesis, Energy Department, Politecnico di Torino. https://ses.jrc.ec.europa.eu/electricity-security, accessed 10th October 2024.

Glaessgen, E. H. and D. S. Stargel. 2012. "The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles." Document No. 20120008178, Langley Research Center (LaRC).

Golnas, A. 2013. "PV System Reliability: An Operator's Perspective." *IEEE Journal of Photovoltaics* 3(1): 416–21.

Grieves, M. and J. Vickers. 2017. "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems." In *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches*, edited by F. Kahlen, S. Flumerfelt, and A. Alves, 85–113. Cham: Springer International Publishing.

Grimmeisen, P., A. Wortmann, and A. Morozov. 2022. "Case Study on Automated and Continuous Reliability Assessment of Software-Defined Manufacturing Based on Digital Twins." In *Proceedings of the 25th International Conference on Model Driven Engineering Languages and Systems: Companion Proceedings (MODELS '22)*, October 23rd-28th, New York, NY, USA, 511–18.

Hou, K., P. Tang, Z. Liu, H. Jia, and L. Zhu. 2021. "Reliability Assessment of Power Systems with High Renewable Energy Penetration Using Shadow Price and Impact Increment Methods." *Frontiers in Energy Research* 9:635071.

Jafari, M., A. Kavousi-Fard, T. Chen, and M. Karimi. 2023. "A Review on Digital Twin Technology in Smart Grid, Transportation System and Smart City: Challenges and Future." *IEEE Access* 11:17471–84.

Johnson, S. C., D. J. Papageorgiou, D. S. Mallapragada, T. A. Deetjen, J. D. Rhodes, and M. E. Webber. 2019. "Evaluating Rotational Inertia as a Component of Grid Reliability with High Penetrations of Variable Renewable Energy." *Energy* 180:258–71.

Kasper, L., F. Birkelbach, P. Schwarzmayr, G. Steindl, D. Ramsauer, and R. Hofmann. 2022. "Toward a Practical Digital Twin Platform Tailored to the Requirements of Industrial Energy Systems." *Applied Sciences* 12(14): 6981.

Lazarova-Molnar, S. 2025. "A Vision for Advancing Digital Twins Intelligence: Key Insights and Lessons from Decades of Research and Experience with Simulation." In *Proceedings of the 14th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2024)*, July 10th-12th, Dijon, France, 5–10.

Lazarova-Molnar, S., P. Niloofar, and G. K. Barta. 2020. "Data-Driven Fault Tree Modeling for Reliability Assessment of Cyber-Physical Systems." In *2020 Winter Simulation Conference (WSC)*, 2719–30 https://doi.org/10.1109/WSC48552.2020.9383882.

Li, Q. and Y. He. 2021. "An Overview of Digital Twin Concept for Key Components of Renewable Energy Systems." *International Journal of Robotics and Automation Technology* 8:29–47.

Liu, W., Z. Lai, K. Bacsa, and E. Chatzi. 2022. "Physics-Guided Deep Markov Models for Learning Nonlinear Dynamical Systems with Uncertainty." *Mechanical Systems and Signal Processing* 178:109276.

Mchirgui, N., N. Quadar, H. Kraiem, and A. Lakhssassi. 2024. "The Applications and Challenges of Digital Twin Technology in Smart Grids: A Comprehensive Review." *Applied Sciences* 14(23): 10933.

Mostafa, O. 2025. "Python for Fault Tree Extraction (PyFTE) GitHub Repository." https://github.com/omarstfa/PyFTE, accessed 11th April 2025.

Nauta, M., D. Bucur, and M. Stoelinga. 2018. "LIFT: Learning Fault Trees from Observational Data." In *Quantitative Evaluation of Systems*, edited by A. McIver and A. Horvath, 306–22. Cham: Springer International Publishing.

Niloofar, P. and S. Lazarova-Molnar. 2023. "Data-Driven Extraction and Analysis of Repairable Fault Trees from Time Series Data." *Expert Systems with Applications* 215:119345.

Niu, M., N. Z. Xu, X. Kong, H. T. Ngin, Y. Y. Ge, J. S. Liu, and Y. T. Liu. 2021. "Reliability Importance of Renewable Energy Sources to Overall Generating Systems." *IEEE Access* 9:20450–59.

Palensky, P., M. Cvetkovic, D. Gusain, and A. Joseph. 2022. "Digital Twins and Their Use in Future Power Systems." *Digital Twin* 1:4.

Pan, H., Z. Dou, Y. Cai, W. Li, X. Lei, and D. Han. 2020. "Digital Twin and Its Application in Power System." In *2020 5th International Conference on Power and Renewable Energy (ICPRE)*, September 12th-14th, Shanghai, China, 21–26.

Sifat, Md. M. H., S. M. Choudhury, S. K. Das, Md. H. Ahamed, S. M. Muyeen, Md. M. Hasan *et al.* 2023. "Towards Electric Digital Twin Grid: Technology and Framework Review." *Energy and AI* 11:100213.

Sonawane, P. R., S. Bhandari, R. B. Patil, and S. Al-Dahidi. 2023. "Reliability and Criticality Analysis of a Large-Scale Solar Photovoltaic System Using Fault Tree Analysis Approach." *Sustainability* 15(5):4609.

Song, Z., C. M. Hackl, A. Anand, A. Thommessen, J. Petzschmann, O. Kamel *et al.* 2023. "Digital Twins for the Future Power System: An Overview and a Future Perspective." *Sustainability* 15(6):5259.

Trivedi, K. S. and A. Bobbio. 2017. *Reliability and Availability Engineering: Modeling, Analysis, and Applications*. 1st ed. Cambridge: Cambridge University Press.

Tuinema, B. W., J. L. Rueda Torres, A. I. Stefanov, F. M. Gonzalez-Longatt, and M. A. M. M. Van Der Meijden. 2020. *Probabilistic Reliability Analysis of Power Systems: A Student's Introduction*. Cham: Springer International Publishing.

Verkuil, B., C. E. Budde, and D. Bucur. 2022. "Automated Fault Tree Learning from Continuous-Valued Sensor Data: A Case Study on Domestic Heaters." *International Journal of Prognostics and Health Management* 13(2).

Yu, W., P. Patros, B. Young, E. Klinac, and T. G. Walmsley. 2022. "Energy Digital Twin Technology for Industrial Energy Management: Classification, Challenges and Future." *Renewable and Sustainable Energy Reviews* 161:112407.

AUTHOR BIOGRAPHIES

OMAR MOSTAFA is is a Ph.D. student at the Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, Germany. His research focuses on the application of Digital Twins to enhance the reliability of energy systems, with a particular focus on the influence of renewable energy sources. omar.mostafa@kit.edu.

SANJA LAZAROVA-MOLNAR is a Professor at both the Karlsruhe Institute of Technology and the University of Southern Denmark. Her research focuses on data-driven simulation, Digital Twins, and cyber-physical systems modeling, with an emphasis on reliability and energy efficiency. She develops advanced methodologies to optimize complex systems and leads several European and national projects in these areas. Prof. Lazarova-Molnar holds leadership roles in IEEE and The Society for Modeling & Simulation International (SCS), where she currently serves as SCS Representative to the Winter Simulation Conference (WSC) Board of Directors. She was Proceedings Editor for WSC in 2019 and 2020 and serves as Associate Editor for *SIMULATION: Transactions of The Society for Modeling and Simulation International*. Her email address is lazarova-molnar@kit.edu.