



### *D1.3. Document on uncertainty representation methods in data driven optimisation models*

Authors: Jean-François Toubeau

This deliverable is based on the following publication:

*Toubeau, J.-F. (2026). Forecasting Techniques for Ancillary Services Provision, Chapter 7 of the Book 'Grid services from wind turbines', to be published in 2026.*



# Table of Contents

<b>Table of Contents</b> .....	2
1. <b>Purpose within the project</b> .....	3
2. <b>Methodological contribution</b> .....	3
2.1. Forecasting as a tool for decision-making .....	3
2.2. Evolution from Statistical Models to Machine Learning .....	4
<b>Statistical Models</b> .....	4
<b>Machine Learning Models</b> .....	4
<b>Ensemble Learning Approaches</b> .....	4
2.3. Main Results and Perspectives .....	4
2.4. Concluding Remarks .....	6

# 1. Purpose within the project

The increasing penetration of renewable energy sources introduces substantial uncertainty into power system operation. Wind and solar generation are inherently variable and difficult to predict accurately, creating challenges for congestion management, reserve procurement, balancing operations, and security assessment. Within the DISCRETE project, uncertainty representation is therefore a central research topic because operational decisions must increasingly be taken under uncertain future system conditions.

The forecasting work addresses this challenge by investigating advanced data-driven methods capable of predicting uncertain variables relevant for power system operation. Rather than directly solving optimisation problems, forecasting models provide quantitative information about future system states that can subsequently be integrated into operational planning and optimisation frameworks.

The role of forecasting within DISCRETE is therefore twofold:

1. **Improving uncertainty quantification**, by generating accurate predictions of renewable generation and ancillary service requirements.
2. **Providing uncertainty-aware inputs to optimisation models**, enabling more informed preventive and corrective operational decisions.

Forecasting thus constitutes an essential first layer in the overall DISCRETE framework. While later work packages focus on learning-based optimisation and risk-aware decision support, these methods fundamentally depend on reliable representations of future uncertainty.

Aspect	Summary
Main objective	Improve representation of uncertainty in renewable-dominated power systems
Methodological focus	Statistical learning, machine learning and ensemble forecasting
Operational value	Improved reserve sizing, balancing, congestion management and market participation

# 2. Methodological contribution

## 2.1. Forecasting as a tool for decision-making

The book chapter emphasizes that operational planning is fundamentally a sequential decision-making problem. System operators must determine actions for future time periods before the actual system state is known. Forecasting techniques therefore aim to predict future values of uncertain variables over the entire planning horizon. These predictions become key inputs for reserve capacity procurement, congestion management, balancing operations, market bidding strategies and security assessment procedures.

The forecasting task is formulated as a supervised learning problem where historical observations, weather information and operational data are used to learn relationships between past and future system states.

An important distinction is made between:

- **Deterministic forecasting**, which provides a single expected future trajectory;
- **Probabilistic forecasting**, which explicitly characterizes uncertainty through probability distributions, prediction intervals or scenarios.

From the perspective of DISCRETE, probabilistic forecasting is particularly valuable because it provides information that can be directly incorporated into uncertainty-aware optimisation models.



## 2.2. Evolution from Statistical Models to Machine Learning

The book chapter reviews the main families of forecasting methodologies and their respective advantages and limitations.

### Statistical Models

Traditional forecasting methods rely on predefined mathematical structures and assumptions regarding the statistical properties of the data. Examples include ARMA models, ARIMA models, seasonal ARMA models and exponential smoothing approaches.

These methods offer several advantages such as strong interpretability, limited computational requirements, and well-understood theoretical foundations.

However, they struggle to accurately represent highly nonlinear behaviours and complex interactions that characterize renewable generation processes.

As renewable-dominated systems become increasingly complex, the limitations of purely statistical approaches become more apparent.

### Machine Learning Models

Machine learning techniques overcome many of these limitations by learning complex relationships directly from data. The document highlights three major families.

First, **tree-based models** partition the feature space into regions associated with different predictions. Representative examples include decision trees, random forests and Gradient Boosting Decision Trees. These techniques are particularly attractive because they handle large feature spaces efficiently, capture nonlinear relationships, provide relatively robust performance and they require limited feature engineering.

Second, **neural networks** constitute highly flexible universal function approximators capable of learning complex nonlinear relationships. Their principal advantage lies in their ability to extract complex hidden structures from large datasets while automatically adapting model complexity to the problem.

These models have become increasingly attractive due to the growing availability of operational data, along with the improved computational resources.

### Ensemble Learning Approaches

Instead of relying on a single model, ensemble techniques combine multiple forecasting methods to exploit their complementary strengths.

The main benefits include improved accuracy, increased robustness, reduced prediction variance, and better generalization capabilities.

Ensemble approaches are particularly relevant for uncertainty representation because they naturally provide multiple plausible future trajectories that can be transformed into probabilistic forecasts or operational scenarios.

For optimisation applications, such ensembles offer a convenient mechanism for generating uncertainty scenarios while maintaining high forecasting quality.

## 2.3. Main Results and Perspectives

A major conclusion of the work is that forecasting should no longer be viewed as an isolated prediction task. Instead, forecasting models must increasingly be designed in conjunction with the downstream optimisation problems they support. It is essential to jointly consider the prediction task and the subsequent decision-making. Indeed, the way we represent prediction uncertainties (e.g., through intervals, full distribution or scenarios) must comply with the characteristics of the downstream decision-making problem. In the following Table, we summarize the four main techniques for optimization under uncertainty. For each technique, we briefly outline how the uncertain variables have to be represented, what is the main goal of the decision-maker along with its risk-attitude, what is the scalability of the resulting model and a short discussion over its main limitation

Characteristics of the main frameworks for optimization under uncertainties: the way we characterize uncertainties is strongly intertwined with the choice of optimization model.

	Scenario-based programming	Robust Optimization	Chance-constrained programming	Distributionally Robust Optimization
Uncertainty	Discrete set of scenarios	Continuous uncertainty set	Distribution shape	Family of distributions close to a « nominal » (ambiguity set)
Goal	Optimize expected value over all scenarios	Optimize for the worst-case value of uncertainty and be feasible for all cases	Optimize objective and be feasible within a given level of confidence	Optimize objective for the worst-case distribution
Risk (reliability)	Risk-neutral, quality depends on scenarios	Risk-averse	Controllable	Controllable via size of ambiguity set and use of chance constraints
Scalability	Hard	Easy	Easy	Hard
Critique	Too risky (bad outcomes in extreme cases)	Too conservative (bad outcomes in mean scenarios)	Assumption on the distribution shape	Need only partial information on the uncertainty

In complement, several future research directions are identified.

**Improved Spatio-Temporal Modelling**

Future forecasting methods should better capture dependencies between geographically distributed renewable assets. Such models would provide more realistic representations of system-wide uncertainty.

**Dynamic Model Recalibration**

Power system dynamics evolve continuously due to changing weather conditions, seasonal effects and market structures. Forecasting models should therefore adapt dynamically to maintain prediction accuracy over time.

**Data Sharing and Collaborative Learning**

The document highlights the growing importance of collaborative learning frameworks that allow multiple stakeholders to leverage complementary datasets while preserving data privacy.

**Explainable Forecasting**

As machine learning models become increasingly complex, improving interpretability becomes essential for operational deployment and operator trust.

**Forecasting-Informed Optimisation**

Perhaps the most important perspective for DISCRETE is the progressive integration of forecasting and optimisation. Rather than treating prediction and decision-making as separate tasks, future frameworks should jointly optimize forecasting quality and operational performance.

This vision aligns directly with the DISCRETE objective of developing risk-aware, data-driven operational tools where uncertainty representation, forecasting and optimisation become tightly integrated components of a unified decision-support framework.



## 2.4. Concluding Remarks

The forecasting work demonstrates that uncertainty representation in modern power systems increasingly relies on advanced data-driven methodologies. While classical statistical methods remain valuable for their interpretability and simplicity, machine learning and ensemble techniques provide superior capabilities for capturing the nonlinear and stochastic nature of renewable generation.

Within DISCRETE, forecasting constitutes the first step towards uncertainty-aware operation. By generating reliable predictive information and probabilistic representations of future system states, forecasting models provide the essential foundation upon which advanced optimisation and decision-support methodologies can be built.