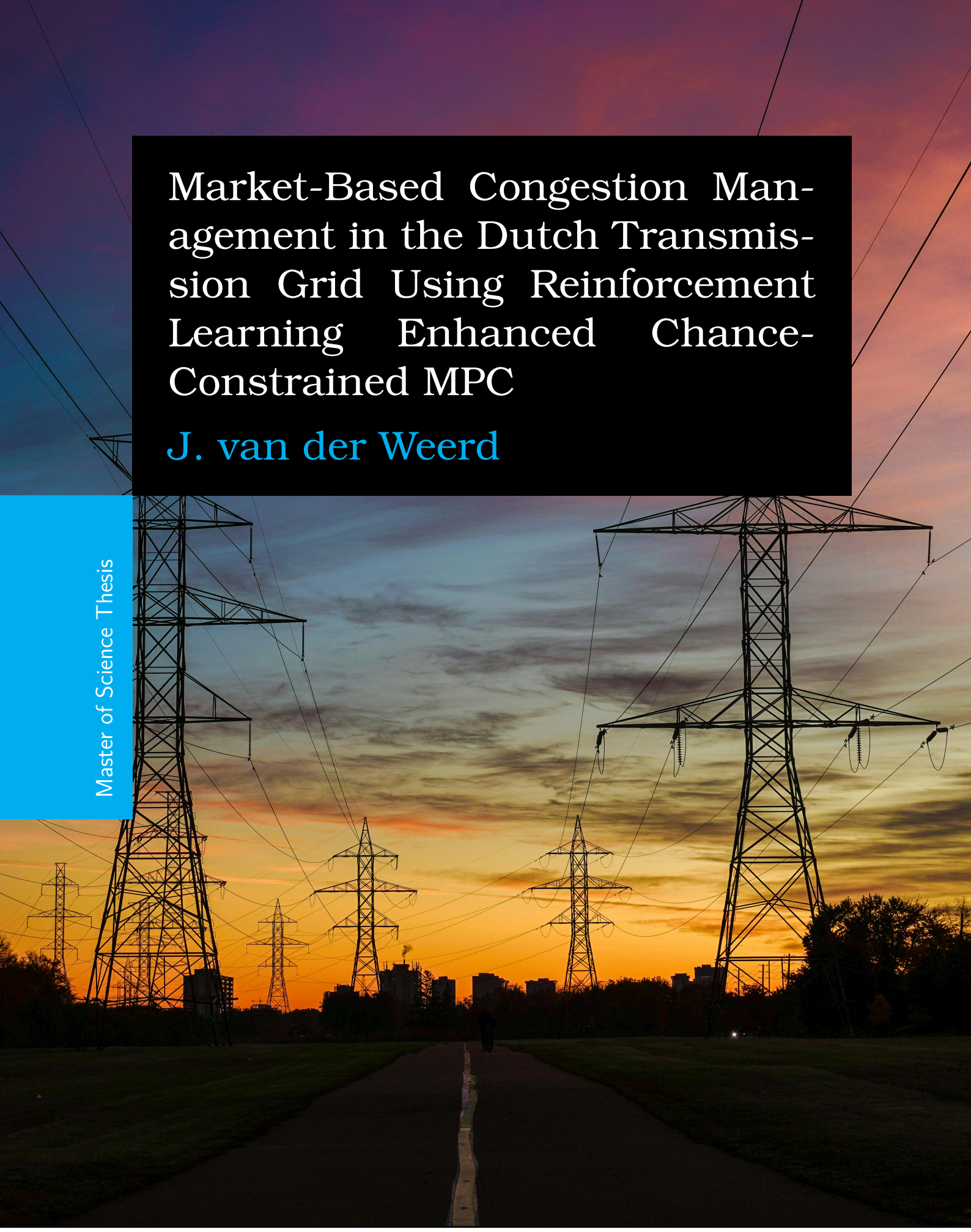
Market-Based Congestion Management in the Dutch Transmission Grid Using Reinforcement Learning Enhanced Chance-Constrained MPC

J. van der Weerd

Master of Science Thesis



TU Delft Delft University of Technology

Delft Center for Systems and Control

Market-Based Congestion Management in the Dutch Transmission Grid Using Reinforcement Learning Enhanced Chance-Constrained MPC

Master of Science Thesis

For the degree of Master of Science in Systems and Control at Delft University of Technology

J. van der Weerd

November 7, 2025

Faculty of Mechanical Engineering (ME) · Delft University of Technology

The work in this thesis was supported by Energie Data Services Nederland (EDNS) and Grid Operators Platform for AnCillary Services (GOPACS). Their cooperation is hereby gratefully acknowledged.

Abstract

Driven by the rapid integration of Renewable Energy Sources (RESs) and the growing electrification of transport, heating, and industry, the Dutch power grid is being fundamentally reshaped. While essential for meeting climate goals, these developments introduce significant operational challenges, including higher uncertainty in power production and congestion risks. Existing approaches for Congestion Management (CM) often neglect the stochastic nature of RESs generation, rely on simplified network representations, or overlook real-world market constraints.

This thesis addresses these gaps by formulating the Dutch market-based CM problem as a Chance-Constrained Model Predictive Control (CC-MPC) framework. A linearized model of the Dutch high-voltage network is employed within a CC-MPC scheme that incorporates flexibility offers through integer decision variables. Uncertainty in RESs generation is captured using an Seasonal AutoRegressive Integrated Moving-Average (SARIMA) model for each production region in the network, enabling a probabilistic treatment of forecast errors. To mitigate conservatism in the chance constraints, a Reinforcement Learning (RL) approach is introduced to adaptively tune the uncertainty model. The resulting stochastic disturbance trajectories are used in a sampling-based approximation of the CC-MPC, optimising congestion mitigation decisions under uncertainty.

The proposed methodology is validated using real-world data from the Dutch energy data exchange platform Energie Data Services Nederland (EDSN), including operational data from Grid Operators Platform for AnCillary Services (GOPACS), the national CM platform. Results demonstrate that the RL-enhanced CC-MPC achieves improved constraint satisfaction compared to other methods. Overall, this work contributes to the current literature by developing a rigorous framework for market-based CM under uncertainty, aimed at ensuring the reliable and cost-effective operation of future renewable-dominated power systems.

Table of Contents

Preface vii

Acknowledgements ix

1 Introduction 1
1-1 Background . . . 1
1-2 Problem description and research question . . . 2
1-3 Structure of this report . . . 3

2 Related works and background information 5
2-1 Relevant background information . . . 5
2-1-1 Operational timescales of power grids . . . 5
2-1-2 Dutch power markets . . . 6
2-1-3 Dutch power grid . . . 7
2-2 Relevant background theory . . . 8
2-2-1 The Alternating Current (AC) Power flow model . . . 8
2-2-2 AutoRegressive Moving Average models . . . 8
2-2-3 Model Predictive Control . . . 10
2-2-4 Reinforcement learning . . . 12
2-3 Related Works . . . 14
2-3-1 Power Flow Models . . . 14
2-3-2 Market-Based CM Methods . . . 14
2-3-3 Control Methodologies for CM . . . 15
2-3-4 Uncertainty Treatment in CM . . . 16
2-3-5 Overview . . . 16
2-4 Summary . . . 16

3 Data usage and uncertainty **19**
3-1 Data description . . . 19
3-2 Data analysis and processing . . . 19
3-2-1 Raw Data Analysis . . . 20
3-2-2 Stationarity Assessment, Differencing, Seasonality Identification . . . 23
3-2-3 Autocorrelation Analysis for Model Order Selection . . . 23
3-2-4 Model Selection using correct Akaike information criterion . . . 24
3-3 Model validation . . . 24
3-4 Summary . . . 28

4 Modelling **29**
4-1 Linearised dynamical transmission network model . . . 29
4-1-1 Limits . . . 31
4-2 Market model . . . 31
4-2-1 Balancing requirement . . . 31
4-2-2 Profile offers . . . 32
4-2-3 Flex-time offers . . . 33
4-2-4 Cost . . . 35
4-3 Summary . . . 35

5 Control methods **37**
5-1 Algorithmic Greedy Matching approach for CM . . . 37
5-2 CM as an Model Predictive Control (MPC) problem . . . 38
5-3 CM as an CC-MPC problem . . . 41
5-4 CM as an CC-MPC-RL problem . . . 42
5-5 Summary . . . 51

6 Case study and results **53**
6-1 Simulation and experimental setup . . . 53
6-1-1 Data usage . . . 54
6-1-2 Offer generation & analysis . . . 55
6-1-3 Case Study — 2024-08-20 . . . 56
6-2 Impact of Prediction Quality on Model Performance . . . 59
6-3 Impact of risk parameter . . . 60
6-4 Impact of RL-Enhancement . . . 60
6-5 Comparison between methods . . . 61
6-6 Summary . . . 63

7 Conclusion, Discussion and Future work **65**
7-1 Conclusion . . . 65
7-2 Discussion . . . 66
7-3 Contributions and Future Work . . . 67
7-3-1 Contributions . . . 67
7-3-2 Future Work . . . 67

A Derivation of the Linearised Model **69**

B Paper style thesis **75**

Bibliography **89**

Glossary **95**

List of Acronyms 95

List of Symbols 95

Preface

This document forms part of my Master of Science graduation thesis for the Systems and Control program at Delft University of Technology.

The idea of conducting my research on CM in the Dutch electricity market originated from my working experience at an energy supplier start-up, where I was first exposed to the operational and economic challenges of energy trading and flexibility management.

This experience sparked my interest in understanding how control strategies and market mechanisms could be combined to create a more efficient and resilient energy network. The thesis represents the culmination of that curiosity and an opportunity to contribute, in a small way, to the ongoing energy transition in the Netherlands.

Throughout the project, I have had the privilege of combining theoretical insights from control engineering with practical challenges from the energy sector. The collaboration with EDSN allowed me to work with real-world data and gain valuable experience in applying academic research to industrial practice.

I hope this work provides useful insights for both researchers and practitioners working at the intersection of power systems, data science, and market design.

Delft, University of Technology — J. van der Weerd
November 7, 2025

Acknowledgements

I would like to thank Prof. dr. ir. B.H.K. De Schutter for the opportunity to conduct this research within his Systems and Control group. His support, particularly the freedom to define my own thesis topic in collaboration with a company, was greatly appreciated.

I would also like to express my sincere gratitude to my academic supervisors, Ir. F. Cordiano and Ir. A. Riccardi, for their invaluable guidance, critical feedback, and continuous support throughout the development of this thesis. Their expertise greatly enhanced both the technical depth and academic quality of my work.

Furthermore, I wish to thank my supervisors from EDSN and GOPACS, Dr. ir. H.M. de Jong, and P. Krootjes for their constructive support and for their guidance within EDSN and GOPACS. The collaboration with EDSNand GOPACS has been both inspiring and educational, offering a valuable perspective on the challenges faced by grid operators in practice.

A special thanks goes to the rest of my colleagues at GOPACS for their warm welcome, good conversations, and genuine interest in my work. Their encouragement and humour made the experience all the more enjoyable.

Chapter 1

Introduction

1-1 Background

In 2015, the Paris Agreement established a commitments of governments across the globe to limit the rise in average global temperature, requiring a rapid reduction in greenhouse gas emissions [1]. One of the key technologies for achieving this goal are RESs such as wind and solar power. Governments worldwide have pledged to triple renewable energy capacity by 2030 [2], and recent trends already show a steep increase in installed capacity.

Simultaneously, the electrification of transport, heating, and industrial sectors is essential, further increasing electricity demand [3]. This transition is driven not only by the need to achieve climate goals but also from a strategic geopolitical standpoint, which necessitates a shift towards energy autonomy [4]. These factors contribute to the expected total increase in electricity demand by 2030 of 60%. To accommodate for the extra demand big investments for expanding and modernising the transmission and distribution grid are necessary and the European commission puts the overall investments around 70 billion euros per year until 2050 [5].

This ambitious shift toward increased electrification, while vital, presents numerous challenges for grid operators. Today's power systems, designed and implemented decades ago, operate under outdated assumptions such as unidirectional power flow and centralized power production. The challenge is two-fold: first, as renewable energy sources like wind and solar become more prevalent, the intermittent nature of these power sources increases the demand for balancing on the grid; second, the traditional approaches that mainly use conventional generators become less effective as their share of total energy requirement decreases. As a consequence, congestion issues become more pronounced, potentially causing voltage violations or thermal overloading of network components [6]. Excessive loading can lead to overheating, voltage instability, or, in extreme cases, even result in outages if not managed properly [7]. This evolution calls for innovative methods that can cope with reduced flexibility from conventional plants, increased distribution of generation, and greater uncertainty in the network.

The importance of efficient CM extends beyond purely technical considerations as detailed above. Economically, congestion leads to high remedial costs, such as re-dispatch or curtailment of renewable generation. For example, in 2023 alone, remedial CM measures in the European power grid cost €4.26 billion [8]. Environmentally, curtailing RES undermines decarbonization efforts and slows progress toward climate targets. Socially, rising costs and reliability concerns hinder industrial expansion and delay the benefits of electrification, i.e., higher energy efficiency and lower energy prices [3], for consumers.

1-2 Problem description and research question

The CM problem is complex, driven by the growing uncertainty in power system operation and the limited range of control options available to transmission system operators. The growing integration of RESs, coupled with a declining share of conventional generation in the overall energy mix, substantially increases variability and uncertainty in both supply and demand. The declining generation share of conventional units also limits the available capacity for remedial actions. Consequently, alternative sources of flexibility from market participants must be utilized. At the same time, market rules and regulatory frameworks restrict the set of actions network operators can take to alleviate congestion. Consequently, ensuring secure and efficient grid operation has become an increasingly challenging task, demanding advanced decision-making methods capable of operating under uncertainty.

Despite the growing importance of CM, existing operational practices often rely on rule-based or heuristic approaches. On the other side, more advanced optimization-based methods are a promising paradigm to tackle the CM problem, as they allow to systematically handle CM and market constraints. Among these, MPC emerges as a promising alternative because it allows explicit incorporation of system constraints and forecasts of future conditions. By optimising remedial actions over a prediction horizon and continuously updating decisions as new data become available, MPC can enhance security, improve economic efficiency, and adapt control actions to evolving system conditions in real time. However, a key limitation of MPC is systematically handling uncertainty, such as forecasting errors in power demand or RESs generation. To address the time-varying nature of these uncertainties, stochastic models (e.g., ARMA) are needed to incorporate probabilistic information directly into the optimization via chance constraints. Since such uncertainty models are necessarily approximate, it remains challenging to select appropriate MPC hyperparameters under changing conditions. To further enhance adaptability and robustness, RL techniques can be integrated to refine these hyperparameters online by learning from past data, improving control performance.

To address the outlined challenges, the proposed approach integrates these strategies into a unified, predictive control framework for CM. The scope and objectives of this work are therefore captured in the following research questions:

How can an MPC-based CM strategy, enhanced with RL, be implemented to manage congestion under uncertainty for the Dutch transmission grid using real-world data?

The research question can be subdivided into the following sub-questions:

- How can the Dutch high-voltage transmission grid be modelled as an MPC problem for CM?
- How can a the Dutch CM market be formulated as an MPC problem?
- How can real-world data be leveraged to incorporate uncertainty within the proposed MPC CM framework?
- How can an RL strategy be leveraged to increase performance within the proposed MPC CM framework?

1-3 Structure of this report

In this section a short overview of the structure of the rest of this thesis is provided. Chapter 2 offers the necessary background on the main topic and its application context, establishing the foundation for the subsequent analysis. Chapter 3 focuses on the data supplied by EDSN, describing in detail its processing and subsequent use in the study. Chapter 5 introduces various control methods, including the novel approach developed as part of this thesis. These methods are then applied and evaluated in Chapter 6 through a case study, where their performance is compared and discussed to demonstrate the effectiveness of the proposed solution. Finally, Chapter 7 summarizes the findings and offers concluding remarks based on the results obtained.

Chapter 2

Related works and background information

This chapter presents and analyses related works, provides essential background information on key theoretical concepts, and establishes the necessary context for the real-world application of the proposed approach. First, the background information on CM in the Netherlands and the Dutch power grid will be presented. Then, an outline of the theoretical foundation of the proposed methodology will be given, focusing on the core concepts of this thesis: power grid modelling, MPC, AutoRegressive Moving-Average (ARMA) estimation, and RL. Finally, related works will be presented and discussed.

2-1 Relevant background information

This section provides the necessary background knowledge a broad introduction to the architecture of power grids and the Dutch power markets, with a specific focus on the GOPACS, the Dutch market for congestion services. Additionally, it includes a mathematical model of the Dutch high-voltage transmission grid, which will be utilised throughout the thesis.

2-1-1 Operational timescales of power grids

Due to the complex nature of the power grid, system control is divided into multiple timescales to ensure each task remains manageable and effective. The shortest timescales fall under operations, where primary control, such as frequency control, must react within seconds, followed by secondary control that restores nominal frequency within minutes. As the required response becomes slower, the tasks shift toward operational planning. This includes tertiary control, i.e. CM, and other system security and economic dispatch actions that operate on a timescale of minutes to hours. Planning over even longer horizons involves decisions like unit commitment over hours to days, and maintenance scheduling, which spans days to months.

Finally, facility planning addresses long-term system development, from day-ahead market considerations to transmission expansion, which may extend over months to years [9].

These categories and representative tasks are illustrated in Figure 2-1. Faster timescales require more detailed and accurate system models because of the rapid physical dynamics involved. Longer-term planning tasks allow for more simplified modelling approaches, since they address slow-evolving structural and economic changes rather than real-time stability.

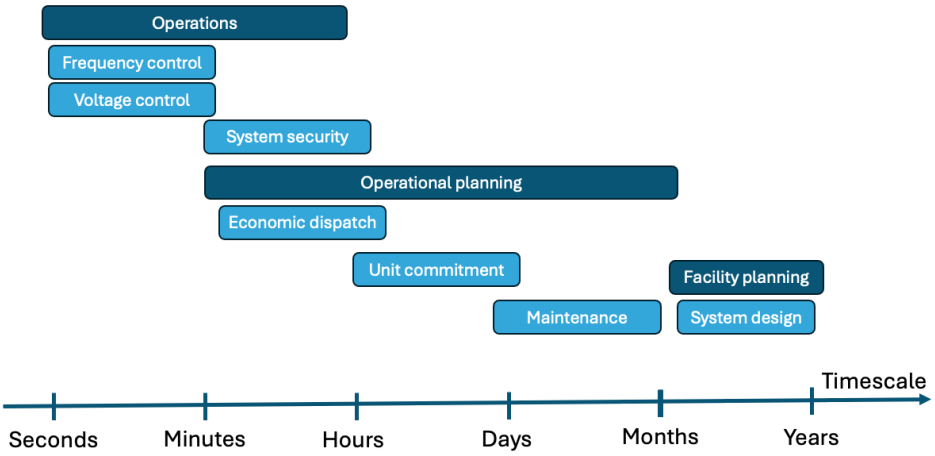


Figure 2-1: Operational timescales for power grids

2-1-2 Dutch power markets

The Dutch electricity market is organized into several layers, each designed to balance supply and demand over different time horizons. These electricity markets range from one day ahead to months or years ahead. Participants can fine-tune their positions in the intra-day market until five minutes before delivery, addressing precise demand and generation needs. Finally, the imbalance market, operated by the transmission system operator TenneT, is used to guarantee a stable system frequency and overall grid stability [10].

CM in the Netherlands is implemented in a market-based manner after market settlement, using the centralized platform GOPACS, which matches flexibility offers submitted by market participants. Transmission and distribution system operators compensate the spread between buy and sell orders: a buy order reflects a downward adjustment in net position (consuming more or generating less), while a sell order reflects an upward adjustment in net position (consuming less or generating more). When the predicted production and consumption patterns show congestion is possible, GOPACS pairs opposing offers to ensure a net-zero power adjustment, preventing any impact on system frequency. The network operator covers the price spread, with positive prices indicating payment by the participant on the buy side or compensation on the sell side. This coordinated mechanism allows congestion to be resolved efficiently without distorting wholesale market prices or jeopardizing grid stability [11].

In this work, the flexibility offers are modelled after the requirements for the new offer structure as specified by the Dutch high-voltage transmission grid operator TenneT [11]. The properties of two different types of orders are

Volume	Time
Partial activation	Start time and duration
Minimum activation	Minimum duration
Quantity	Maximum duration

Table 2-1: Properties different market order types

2-1-3 Dutch power grid

The focus of this thesis is CM in the Dutch high-voltage transmission grid. In this thesis, the high-voltage grid is represented as a undirected mathematical graph, based on the map made by Tennet [12], consisting of nodes and edges in the following way:

$$\begin{aligned} &\mathscr{G} = (\mathscr{N}, \mathscr{E}) \\ &\mathscr{N} := \{0, \ldots, |\mathscr{N}| - 1\} \\ &\mathscr{E} \subseteq \{(n, m) \mid n, m \in \mathscr{N}, n \neq m\} \end{aligned} \tag{2-1}$$

where $\mathcal{G}$ is the undirected graph, $\mathcal{N}$ denotes the set of nodes and $\mathcal{E}$ denotes the set of edges. The number of nodes is represented by $|\mathcal{N}|$. The graphical representation of this network is shown in Figure 2-2. Where the nodes $\{39, 40, 41, 42, 43, 44, 45, 46\}$ represent the medium-voltage regions that connect all consumption and production to the high-voltage grid. The rest of the nodes are high-voltage substations.

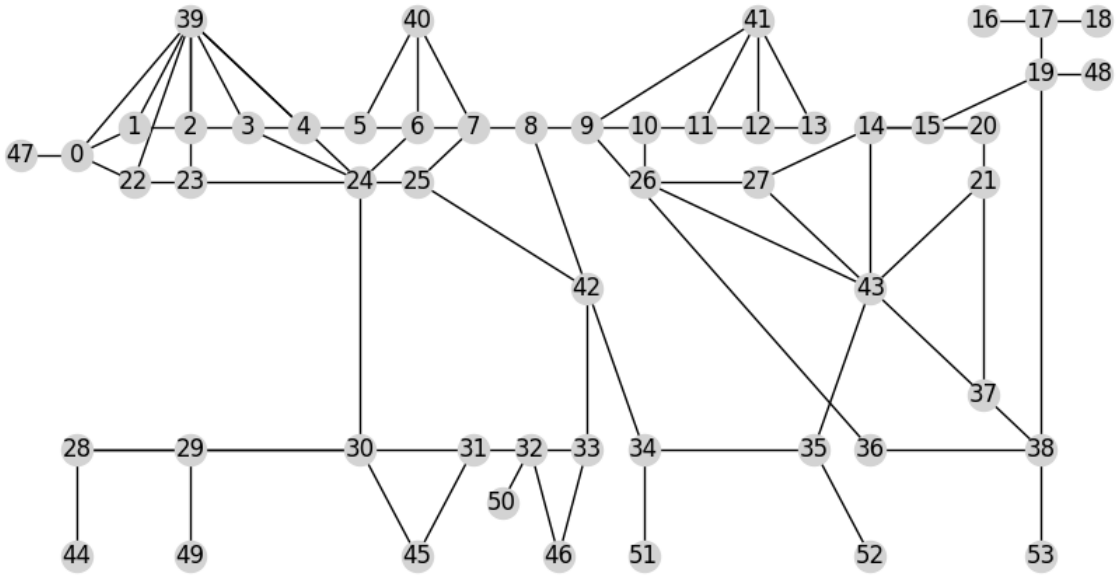


Figure 2-2: Graph of the studied high-voltage power grid, where each node represents a coupling substation and each edge represents a transmission line connection between substations

2-2 Relevant background theory

In this section, the necessary theoretical background for this thesis is presented. It covers four main topics: the power flow model, ARMA models, MPC, and RL.

2-2-1 The AC Power flow model

Many power flow models exist, ranging from highly detailed representations that capture device-level dynamics to simplified linearised formulations. The choice of model typically involves a trade-off between accuracy and computational tractability. In this thesis, the focus is on the behaviour of a transmission network. Since the network under consideration is a balanced three-phase systems, it is sufficient to adopt a single-line representation of the grid [13]. In this model the active and reactive power balance at each node n at time t is expressed as in [14]:

$$\begin{aligned} P^{(n)}(t) = \sum_{m \in \mathcal{N}} \Big(& - v^{(n)}(t) v^{(m)}(t) g^{(n,m)} \cos\left(\theta^{(n)}(t) - \theta^{(m)}(t)\right) \\ & - v^{(n)}(t) v^{(m)}(t) b^{(n,m)} \sin\left(\theta^{(n)}(t) - \theta^{(m)}(t)\right) \Big) \end{aligned} \tag{2-2a}$$

$$\begin{aligned} Q^{(n)}(t) = \sum_{m \in \mathcal{N}} \Big(& - v^{(n)}(t) v^{(m)}(t) b^{(n,m)} \cos\left(\theta^{(n)}(t) - \theta^{(m)}(t)\right) \\ & - v^{(n)}(t) v^{(m)}(t) g^{(n,m)} \sin\left(\theta^{(n)}(t) - \theta^{(m)}(t)\right) \Big), \end{aligned} \tag{2-2b}$$

where $P^{(n)}(t)$ and $Q^{(n)}(t)$ denote the active and reactive power injections at bus n at time t, $v^{(n)}(t)$ and $\theta^{(n)}(t)$ are the bus voltage magnitude and phase angle, and $g^{(n,m)}, b^{(n,m)}$ are the conductance and susceptance of line (n,m), respectively. The active and reactive power flows along each transmission line $(n,m) \in \mathcal{E}$ are described as

$$\begin{aligned} P^{(n,m)}(t) = & (v^{(n)}(t))^2 g^{(n,m)} - v^{(n)}(t) v_t^{(m)} g^{(n,m)} \cos\left(\theta^{(n)}(t) - \theta^{(m)}(t)\right) \\ & - v^{(n)}(t) v^{(m)}(t) b^{(n,m)} \sin\left(\theta^{(n)}(t) - \theta^{(m)}(t)\right) \end{aligned} \tag{2-3a}$$

$$\begin{aligned} Q^{(n,m)}(t) = & -(v^{(n)}(t))^2 b^{(n,m)} + v^{(n)}(t) v^{(m)}(t) b^{(n,m)} \cos\left(\theta^{(n)}(t) - \theta^{(m)}(t)\right) \\ & + v^{(n)}(t) v^{(m)}(t) g^{(n,m)} \sin\left(\theta^{(n)}(t) - \theta^{(m)}(t)\right) \end{aligned} \tag{2-3b}$$

where $P^{(n,m)}(t)$ and $Q^{(n,m)}(t)$ are the active and reactive power flows from bus n to bus m at time t. Together, (2-2) and (2-3) define the AC power flow model that will be used in subsequent sections to analyse the behaviour of the transmission grid.

2-2-2 AutoRegressive Moving Average models

ARMA models are a family of linear time series models designed to describe stationary stochastic processes. Given two parameters $p, q \in \mathbb{N}$, an ARMA model combines two complementary components. The AutoRegressive part captures how the current value of the time series depends on its own past observations—specifically, a linear combination of the previous p values. In contrast, the Moving Average part describes how the current value is

influenced by random shocks or innovations from the past q time steps, representing the effect of unmodelled disturbances or noise that persist over time. Together, these two terms allow the ARMA model to represent both deterministic dependencies on past values and stochastic effects arising from past random disturbances. Following [15], an ARMA(p, q) process for a zero-mean stationary time series $\{y(t)\}$ can be expressed as

$$y(t) = \sum_{i=1}^{p} \phi_i y_{t-i} + e(t) + \sum_{j=1}^{q} \theta_j e_{t-j}, \tag{2-4}$$

where $y(t)$ denotes the value of the time series at time t, y_{t-i} are the lagged observations, and $e(t)$ represents a white noise innovation term with zero mean and constant variance. The coefficients ϕ_i and θ_j correspond to the autoregressive and moving average parameters, respectively, while p and q denote the orders of the AR and MA components.

The formulation in (2-4) can be rewritten in a more compact form using the back-shift operator D, defined such that $D^k y(t) = y_{t-k}$. This operator-based representation simplifies the manipulation of lagged variables. Expressed in this way, the ARMA(p, q) model becomes [16]

$$\begin{aligned}(1 - D\phi_1 - D^2\phi_2 + \ldots + D^p\phi_p)y(t) &= (1 + D\theta_1 + D^2\theta_2 + \ldots + D^q\theta_q)e(t),\\ \phi_p(D)y(t) &= \theta_q(D)e(t),\end{aligned} \tag{2-5}$$

where the autoregressive and moving average polynomials are defined as

$$\phi_p(D) = 1 - D\phi_1 - D^2\phi_2 - \ldots - D^p\phi_p \tag{2-6}$$

and

$$\theta_q(D) = 1 + D\theta_1 + D^2\theta_2 + \ldots + D^q\theta_q. \tag{2-7}$$

Because many real-world time series are non-stationary, a preprocessing step called differencing is often used as a preprocessing step. Differencing subtracts consecutive observations from one another to remove trends. A first order differencing operation looks like

$$(1 - D)y(t) = y(t) - y_{t-1}. \tag{2-8}$$

Incorporating differencing of order d yields the AutoRegressive Integrated Moving-Average (ARIMA)(p, d, q) model [17], which modifies (2-5) to

$$\phi_p(D)(1 - D)^d y(t) = \theta_q(D)e(t), \tag{2-9}$$

where $(1 - D)^d$ denotes the d-th order differencing operator applied to the series.

When the time series exhibits periodic or seasonal behaviour, the model can be further extended to include seasonal components, resulting in the SARIMA$(p, d, q, p_s, d_s, q_s, s)$ [18]. This model includes both non-seasonal and seasonal autoregressive and moving average terms, and can be written as

$$\phi_p(D)(1 - D)^d \Phi(D^s)(1 - D^s)^{d_s} y(t) = \theta_q(D)\Theta_Q(D^s)e(t), \tag{2-10}$$

where $\Phi(D^s)$ and $\Theta_Q(D^s)$ represent the seasonal autoregressive and moving average polynomials with a seasonal lag of s. The parameters p_s, d_s, and q_s correspond to the orders of the seasonal AR, seasonal differencing, and MA components, respectively, while s denotes the

length of the seasonal cycle (for instance, $s = 12$ for monthly data with annual seasonality). The first order seasonal differencing looks like:

$$(1 - D^s)y(t) = y(t) - y_{t-s} \tag{2-11}$$

This formulation allows the model to simultaneously capture both short-term dynamics and long-term seasonal patterns within a single framework.

Model selection for ARMA, ARIMA, or SARIMA models involves determining the optimal orders (p, d, q) and, in the seasonal case, (p_s, d_s, q_s, s). This choice should balance model complexity and predictive accuracy. To evaluate the models, information criteria such as the Bayesian information criterion [19], or Akaike information criterion [20] are commonly used. In practice, these criteria are used in automated algorithms such as the procedure introduced in [21] to evaluate the models found in a systematic search of the parameter space too find the best model order. Once the model structure is chosen, estimation of the model parameters $(\phi_i, \theta_j, \Phi_i, \Theta_j)$ is typically performed via maximum likelihood estimation. The principle of maximum likelihood estimation is to find the parameter values that maximize the likelihood of the observed data under the assumed model.

2-2-3 Model Predictive Control

MPC is a model-based optimal control strategy that determines a sequence of control inputs to steer the state of a system towards a desired reference trajectory. At each time step, an optimisation problem is solved over a finite prediction horizon, balancing performance objectives (encoded in the cost function) with safety and operational requirements (encoded as constraints on the states and control inputs). This is illustrated in Figure 2-3. The solid black line represents the measured system state $x(k)$, while the orange line shows the reference trajectory $r(k)$ that the system should follow. At the current time step, the controller predicts the future evolution of the system over the horizon n_p based on a sequence of candidate control inputs $u(k)$, shown as green dots. The black dotted line represents the predicted state trajectory resulting from these candidate inputs. Among all feasible input sequences, the optimisation selects the one that minimises the cost while satisfying the state constraints. Only the first input of this sequence is applied to the system, after which the procedure is repeated at the next time step using updated measurements. Deterministic MPC is formulated under a deterministic setting, meaning that it assumes perfect knowledge of system parameters and neglects the presence of disturbances or uncertainties [22]. The deterministic MPC optimisation problem for a linear system is typically written as

$$\min_{\substack{x(k+1),\dots x(k+n_\mathrm{p}),\\ u(k+1),\dots,u(k+n_\mathrm{p}-1)}} \quad \sum_{i=1}^{n_\mathrm{p}-1} l(x(k+i), u(k+i)) + l_\mathrm{f}(x(k+n_\mathrm{p})) \tag{2-12a}$$

$$\text{s.t.} \quad x(k+i+1) = Ax(k+i) + Bu(k+i+1) \quad \forall i \in \{0, \dots, n_\mathrm{p}-1\} \tag{2-12b}$$

$$h(x(k+i+1), u(k+i)) \leq 0 \quad \forall i \in \{1, \dots, n_\mathrm{p}-1\} \tag{2-12c}$$

$$x(k+i) \in \mathcal{X} \quad \forall i \in \{1, \dots, n_\mathrm{p}\} \tag{2-12d}$$

$$u(k+i) \in \mathcal{U} \quad \forall i \in \{1, \dots, n_\mathrm{p}-1\}, \tag{2-12e}$$

here $x(k)$ denotes the measured system state at the current time k and n_p is the prediction horizon. The state matrices A, B form the model describing the state evolution;

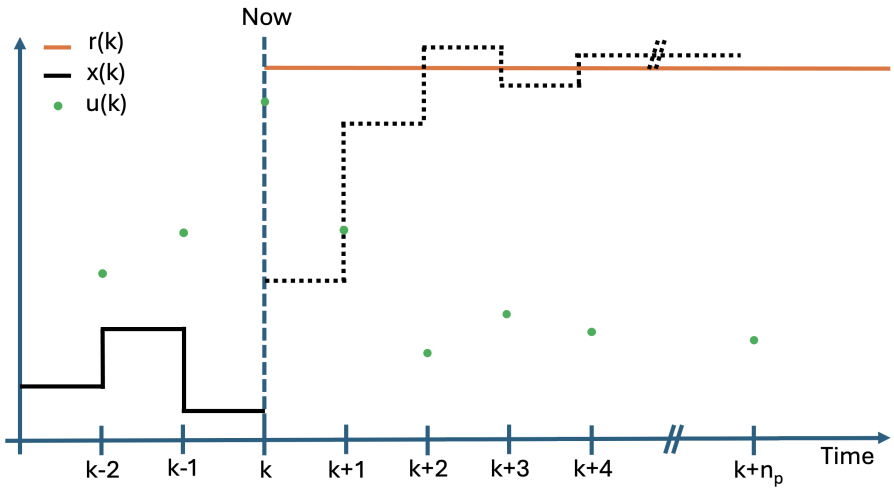


Figure 2-3: Nominal MPC control strategy. The solid black line shows the measured state $x(k)$, the orange line indicates the reference trajectory $r(k)$, and the green dots represent the candidate control inputs $u(k)$ predicted over the horizon n_p.

$h(x(k+i+1), u(k+i))$ encodes the constraints on the states and inputs; and $\mathcal{X} \subseteq \mathbb{R}^n$, $\mathcal{U} \subseteq \mathbb{R}^m$ are the admissible state and control sets. The objective function is composed of two terms: the stage cost $l(x(k+i), u(k+i))$, which penalises state trajectories and control actions along the prediction horizon, and the terminal cost $l_\mathrm{f}(x(k+n_\mathrm{p}))$, which penalises the final predicted state. Despite its effectiveness, deterministic MPC does not explicitly handle uncertainty. However, in practice systems are subject to disturbances, measurement noise, and parameter variations. Incorporating uncertainty into the optimization problem improves decision-making, reliability, and constraint satisfaction, although it generally comes with increased computational complexity. Two main extensions of MPC have been proposed for this purpose: robust MPC and stochastic MPC.

Robust MPC accounts for all possible realisations of uncertainty within a predefined uncertainty set. This yields strong safety guarantees, but often at the expense of performance due to conservatism. In contrast, stochastic MPC relaxes these deterministic safety guarantees into probabilistic constraint satisfaction. Instead of requiring absolute satisfaction of the constraints, it allows for constraint violation with a prescribed probability. This results in less conservative control actions and improved performance, though without absolute safety guarantees [23]. The deterministic formulation in (2-12) can be extended to the stochastic setting as a CC-MPC problem as follows [24]:

$$
\begin{aligned}
\min_{\boldsymbol{x}^{(1)}(k),\ldots,\boldsymbol{x}^{(n_\mathrm{s})}(k),\boldsymbol{u}(k)} \quad & \frac{1}{n_\mathrm{s}} \sum_{\mathrm{s}=1}^{n_\mathrm{s}} \sum_{i=1}^{n_\mathrm{p}-1} l(x^{(\mathrm{s})}(k+i), u(k+i)) + l_\mathrm{f}(x^{(\mathrm{s})}(k+n_\mathrm{p})) \\
\text{s.t.} \quad & x^{(\mathrm{s})}(k+i+1) = Ax^{(\mathrm{s})}(k+i) + Bu^{(\mathrm{s})}(k+i+1) + Dw^{(\mathrm{s})}(k+i+1) \\
& \qquad\qquad \forall i \in \{0,\ldots,n_\mathrm{p}-1\}, \mathrm{s} \in \{1,\ldots,n_\mathrm{s}\} \\
& \mathbb{P}(h(x(k+i+1), u(k+i), d(k+i)) \leq 0, \forall i \in \{1,...,n_p-1\}) \geq \alpha \\
& u(k+i) \in \mathcal{U} \quad \forall i \in \{1,...,n_p-1\}
\end{aligned}
\tag{2-13}
$$

where the disturbance $w(k+i)$ belongs to the uncertainty set $\mathcal{W}$m, $\alpha \in (0,1]$ denotes the minimum probability of constraint satisfaction, and the bold symbols denoted the full trajectory over the prediction horizon. Solving chance-constrained problems is notoriously difficult due to the complexity of the probability constraints, even for relatively simple systems [24]. To enable practical solutions, scenario-based or sampling-based approximations of chance constraints are widely adopted. In these approaches, the probabilistic constraint in (2-13) is replaced by n_s deterministic constraints, one for each sampled scenario. This has two advantages: it transforms the problem into a deterministic program with manageable complexity, and it does not require an explicit knowledge of the uncertainty distribution, provided that a sufficient number of samples can be drawn [23]. Under suitable conditions, the required number of samples can be small, making scenario-based MPC an attractive approach even for large-scale systems [25]. The sample-based approximation of the chance-constrained problem formulated in (2-13) can be written as

$$\begin{aligned}
\min_{\boldsymbol{x}^{(1)}(k),\ldots,\boldsymbol{x}^{(n_\mathrm{s})}(k),\boldsymbol{u}(k)} \quad & \frac{1}{n_\mathrm{s}} \sum_{\mathrm{s}=1}^{n_\mathrm{s}} \sum_{i=1}^{n_\mathrm{p}-1} l(x^{(\mathrm{s})}(k+i), u(k+i)) + l_\mathrm{f}(x^{(\mathrm{s})}(k+n_\mathrm{p})) \\
\text{s.t.} \quad & x^{(\mathrm{s})}(k+i+1) = Ax^{(\mathrm{s})}(k+i) + Bu^{(\mathrm{s})}(k+i+1) + Dw^{(\mathrm{s})}(k+i+1) \\
& \qquad\qquad \forall i \in \{0,\ldots,n_\mathrm{p}-1\}, \mathrm{s} \in \{1,\ldots,n_\mathrm{s}\} \\
& \frac{1}{n_\mathrm{s}} \sum_{\mathrm{s}=1}^{n_\mathrm{s}} \mathbb{1}\left(h(x(k+i+1), u(k+i), d(k+i)) \leq 0, \forall i \in \{1,...,n_p-1\}\right) \geq \alpha \\
& u(k+i) \in \mathcal{U} \quad \forall i \in \{1,...,n_p-1\}
\end{aligned} \tag{2-14}$$

where $\mathbf{x}_k^{(\mathrm{s})}$, denotes the predicted state trajectory under scenario s. Each scenario corresponds to a particular sampled realisation of the disturbance sequence $d^{(\mathrm{s})}(k+i)$, which is drawn from the uncertainty set $\mathcal{D}$. The control input sequence $\boldsymbol{u}_k$ is common to all scenarios, since the controller cannot adapt its decisions to a disturbance before it has been observed. The number of scenarios considered is n_s. The probabilistic constraint in (2-13) is replaced by it sample-based approximation of the joint chance-constraint:

$$\frac{1}{n_\mathrm{s}} \sum_{\mathrm{s}=1}^{n_\mathrm{s}} \mathbb{1}\left(h(x(k+i), u(k+i), d(k+i)) \leq 0, \forall i \in \{1,...,n_p-1\}\right) \geq \alpha, \tag{2-15}$$

which enforces that the fraction of sampled scenarios satisfying the original constraints must be bigger than α. In this way, the chance constraint is approximated through a finite set of deterministic constraints, one for each scenario where $\mathbb{1}$ is defined as in (2-16).

$$\mathbb{1}(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{2-16}$$

2-2-4 Reinforcement learning

RL is one of the three main branches in machine learning, alongside supervised and unsupervised learning. In contrast to supervised learning, RL does not rely on labelled data. Instead; it focuses on learning a policy, a strategy that maps states of the system to actions, with the objective of maximizing the cumulative reward over time.

The RL framework is composed of two main components: the agent and the environment. The agent serves as the decision-maker, while the environment represents the system or process with which the agent interacts. At each step of the learning cycle, the agent observes the state of the environment, selects an action, and subsequently receives both a new state and a reward. This continuous interaction between the agent and the environment is illustrated in Figure 2-4. The environment is often modelled as a Markov decision process. A Markov

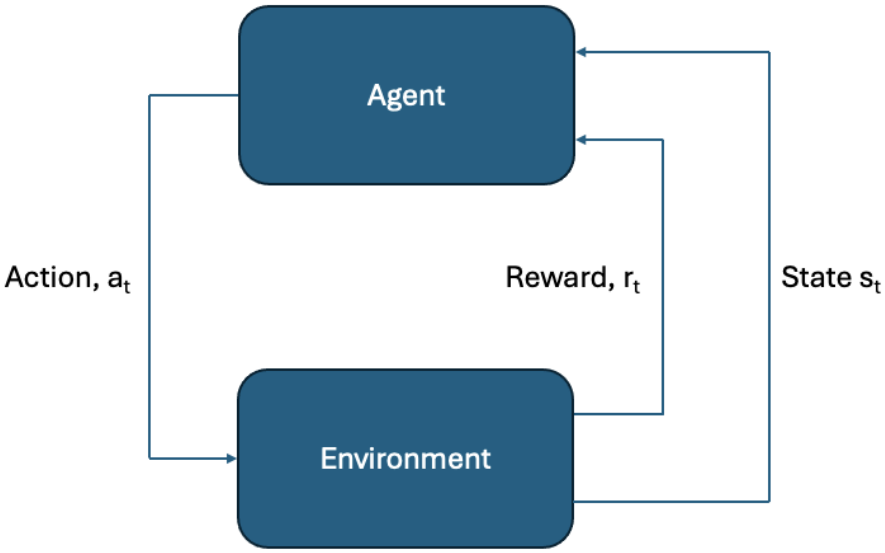


Figure 2-4: Agent-environment interaction in RL

decision process is a framework for sequential decision-making in stochastic settings, defined by states, actions, transition probabilities, and rewards. A Markov decision process can be modelled mathematically as the Markov decision process tuple in the following way [26]

$$\mathcal{M} \sim (\mathcal{S}, \mathcal{A}, \mathbb{P}, R),$$

where $\mathcal{M}$ is the Markov decision process tuple, $\mathcal{S}$ represents the state space, $\mathcal{A}$ represents the action space, $\mathbb{P}$ defines the transition probabilities from the current state $s(t)$ to the next state s_{t+1} under action $a(t)$, and R is the reward function. Note that in this thesis, both s_k and x_k are referred to as states; however, the former denotes the state of the RL agent, while the latter represents the state of the power grid's dynamic model. A valid Markov decision process must satisfy the Markov property, which assumes the process is memoryless. This implies that transition probabilities for the next state depend solely on the current state. Formally, this is expressed as

$$P(s_{k+1} \mid s_k, a_k) = P(s_{k+1} \mid s_k, a_k \ldots, s_0, a_0), \tag{2-18}$$

where $P(\cdot)$ is the transition probability from state $s_k \in \mathcal{S}$ to state $s_{k+1} \in \mathcal{S}$ under action a_k. If the environment can be modelled as an Markov decision process, utilizing only the current state to determine an action is as effective as using the complete history of states.

2-3 Related Works

This section reviews related work on market-based CM relevant to this thesis. The discussion is organized around four main themes: power flow models, market-based CM approaches, control methodologies, and uncertainty treatment in CM. Each topic is examined to highlight current methods, their underlying assumptions, and key differences. The section concludes with a summary of the reviewed literature and a discussion of the research gap addressed by this thesis.

2-3-1 Power Flow Models

Two types of power flow models are commonly applied in CM studies: the AC power flow model and the Direct Current (DC) power flow model.

The AC model provides a detailed and accurate representation of the electrical grid by capturing both real and reactive power flows through non-linear equations based on power balance at each node [27, 28]. It considers voltage magnitudes, voltage angles, conductance, and susceptance for each line, allowing precise estimation of voltages, reactive power flows, and power losses. However, its non-linear nature makes the model computationally intensive and less suitable for large-scale or real-time applications.

This complexity motivates the use of simplified alternatives when high accuracy is not essential. The most common simplification is the DC power flow model, which assumes: no active power losses, no reactive power, small voltage angle differences between nodes, and a uniform voltage magnitude of one per unit at all nodes [29, 30]. These approximations justify the linearisation of the AC equations, allowing faster computation and efficient estimation of active power flows. However, these simplifications limit accuracy. The DC model performs well only when the reactance-to-resistance ratio exceeds four and voltage angle differences remain below about 7°, conditions typically found in high-voltage transmission networks. The flat voltage profile assumption is its main limitation, as it can lead to large errors under varying load or voltage regulation conditions. Therefore, while the DC model offers computational efficiency, its validity depends strongly on the network characteristics and operating conditions [29].

2-3-2 Market-Based CM Methods

Market-based CM can be broadly categorized into ex-ante and ex-post approaches, depending on whether transmission limits are considered during or after market settlement.

In ex-ante markets, such as nodal or zonal markets, transmission constraints are integrated directly into the market-clearing process. In a nodal market, electricity prices are node-specific; when transmission capacity to a node is scarce, the local price increases to reflect congestion costs, and only the highest offers are cleared [31–33]. In a zonal market, multiple nodes are grouped into zones with uniform pricing, simplifying settlement but neglecting congestion within zones [34]. In principle, ex-ante markets should eliminate the need for remedial actions since congestion is priced into the settlement. In contrast, ex-post CM operates after market clearing. The main market is settled as a single zone under the copper plate assumption (i.e., unlimited transmission capacity), and congestion is resolved subsequently [27]. This can be

achieved through non-market mechanisms such as load shedding or re-dispatch, where flexibility is demanded as needed [35]. Some non-market price-based coordination schemes use price signals as control incentives rather than as market-clearing outcomes. For example, [36] employ a price-based coordination mechanism where the network operator sends incentive signals to aggregators to adjust demand and relieve congestion, without actual offering or clearing. Similarly, [37] use cost-based predictive control to coordinate flexibility ex-post. Such approaches mimic market behaviour but do not involve real trading.

In full market-based approaches, flexibility is traded through pool-based markets, bilateral, or multilateral contracts. Re-dispatch objectives may focus on cost minimization or deviation minimization, depending on whether the goal is to reduce re-dispatch costs or deviations from the market schedule. Flexibility offers are typically represented as offers from participants, submitted as fixed prices or contract-based offers [37–40].

2-3-3 Control Methodologies for CM

Existing CM control methods can be broadly categorized into optimization-based and MPC-based methods

Optimization-based methods formulate CM as a mathematical optimization problem that minimizes system costs or maximizes social welfare subject to physical and operational constraints. These approaches provide static solutions assuming accurate system models and perfect information. Depending on the formulation, objectives include social welfare maximization before market clearing [31, 33], or cost/deviation minimization during re-dispatch [37, 41, 42]. Extensions include multi-objective optimization [43], distributed optimization [28, 44], and mixed-integer formulations for both continuous an discrete control decisions [40, 45].

MPC extends traditional optimization by introducing feedback and receding-horizon control. Instead of computing a static, open-loop solution, MPC repeatedly solves a constrained optimization problem over a prediction horizon, applies the first control action, and re-optimizes as new measurements arrive. This allows the controller to anticipate and mitigate future congestion while adapting to evolving grid conditions. Recent studies [39, 46–48] demonstrate MPC's effectiveness for real-time coordination of generators, storage, and curtailment. Robust MPC extensions further improve resilience to forecast errors and renewable fluctuations [49]. However, MPC requires accurate dynamic models and incurs higher computational costs compared to static optimization. Its performance is also sensitive to how uncertainty is represented inside the controller. In CC-MPC, probabilistic feasibility is enforced using generated scenarios from an underlying statistical model of forecast errors. If the uncertainty model is misaligned with real grid behaviour, these scenarios can lead to under- or over estimation of risk. To mitigate this issue, reinforcement learning is used to adaptively tune the variance parameters of the ARMA-based uncertainty model based on the observed system response. Through this online adjustment, the RL agent balances constraint satisfaction and economic efficiency.

2-3-4 Uncertainty Treatment in CM

Uncertainty in CM arises from forecast errors, imperfect system models, and the variability of RESs. As distributed generation and flexible demand increase, managing these uncertainties becomes crucial for reliable and efficient operation.

Traditional optimization approaches address uncertainty through fixed safety margins on constraints, but this is often overly conservative or fails to guarantee feasibility under large deviations [50]. Recent work therefore focuses on explicit uncertainty modelling using data-driven forecasting methods. Probabilistic and machine-learning models are applied to capture forecast errors and temporal dependencies, providing statistical information that can be incorporated directly into control formulations [51–54]. Probabilistic formulations, such as chance-constrained or scenario-based methods, use probability distributions to balance risk and performance, albeit at higher computational cost [55–57]. Robust optimization ensures feasibility for all realizations within bounded uncertainty sets, making it suitable for data scarcity, or safety-critical applications but often yielding conservative solutions [58].

2-3-5 Overview

Most market-based ex-post approaches rely either on deterministic forecasts [36–39] or static optimization methods that lack closed-loop adaptability. Although recent studies incorporate probabilistic information [40, 55], they do not leverage receding-horizon control or real-time learning to handle evolving uncertainties. Additionally, only a limited number of works combine realistic grid modelling with real-world data.

These gaps are addressed by combining statistical forecasting to capture renewable and demand variations, with a market-based CC-MPC incorporating real market restrictions with a receding horizon control method. Finally an RL-agent is then integrated to tune the ARMA model to improve constraint satisfaction and economic efficiency. This integrated approach demonstrates superior performance relative to a greedy selection approach, deterministic MPC and CC-MPC in terms of constraint satisfaction and economic efficiency, while being validated on real-world data and an AC grid model representative of the Dutch transmission network.

2-4 Summary

This chapter establishes the context for market-based CM in the Netherlands and outlines the theoretical tools used in the thesis. It introduces the Dutch market architecture, from futures to imbalance—and explains how CM is performed ex-post via the GOPACS platform, where opposing flexibility offers are paired and the operator covers the price spread. The Dutch high-voltage grid is abstracted as an undirected graph used throughout the work, and offer attributes for flexibility products (timing and volume constraints) are aligned with TenneT's specification. Last, related works and research gaps are discussed, and the main thesis contributions are outlined.

Paper	Power flow model	Market-based approach	Uncertainty treatment	Control methodology	Data type & grid model
[36]	AC	Ex-post, non-market-based, price coordination	Deterministic forecast	Optimization	Simulated data + benchmark network
[37]	AC	Ex-post, non-market-based, price coordination	Deterministic forecast	MPC	Simulated data + benchmark network
[38]	DC	Ex-post, market-based, pool market	Deterministic forecast	Optimization	Simulated data + benchmark network
[39]	DC	Ex-post, market-based, pool market	Deterministic forecast	Optimization	Simulated data + benchmark network
[40]	AC	Ex-post, market-based, pool market	Statistical	Optimization	Real-world data + real grid model
[55]	DC	Ex-post, market-based, pool market	Statistical	Chance-constrained optimization	Simulated data + benchmark network
This work	AC	Ex-post, market-based, pool market	Statistical	Chance-constrained MPC	Real-world data + real grid model

Table 2-2: Summary of related work on market-based CM

Chapter 3

Data usage and uncertainty

In this chapter, the data and statistical model used for making the disturbance prediction are presented. First, in Section 3-1, the structure of the used data set is described. Then, in Section 3-2, the process of constructing a statistical forecasting model from the raw time series is described, starting with data analysis that includes seasonal identification and stationarity testing, and concluding with the selection of the SARIMA model. The resulting model is evaluated through residual analysis and out-of-sample prediction in Section 3-3, forming the basis for uncertainty representation in the control framework introduced in the following chapters.

3-1 Data description

In the model developed for this thesis, eight areas are considered, each featuring both electricity consumption and production. For each area, the available data are divided into two categories: connections that solely represent consumption and all other connections. The data have been provided by EDSN.

It should be noted that large-scale grid connections exceeding 60 MW are excluded from the dataset, as they are directly connected to the transmission grid and therefore not included in the EDSN data. This exclusion encompasses all conventional power plants. Nevertheless, because the overall power system must remain balanced, the total generation across the system must equal the net demand. This relationship enables the total generation to be estimated and subsequently allocated to the corresponding generator nodes.

3-2 Data analysis and processing

This section develops the statistical model used to make the disturbance forecasts in the proposed CC-MPC framework. The workflow follows well-established guidelines for time-series analysis and SARIMA modelling [21, 59].

Ensuring stationarity is essential for ARMA-type modelling, since if the underlying process changes the models parameters on fitted on past data become biased or meaningless. Therefore, the process begins with an exploration of the raw measurement data to understand its key characteristics, such as the existence of trends or patterns, indicating non-stationarity. Trends and/or patterns can be removed using differencing. To characterize periodic behaviours, a fast Fourier transform is applied, allowing dominant seasonal frequencies to be identified and helping determine the appropriate seasonal period for seasonal differencing [60]. To validate the differencing, statistical tests can be performed to check for stationarity using a unit root test. Two commonly used test are The augmented Dickey–Fuller and Kwiatkowski–Phillips–Schmidt–Shin tests [59,60]. The augmented Dickey–Fuller test assumes that the series is non-stationary, whereas the Kwiatkowski–Phillips–Schmidt–Shin test assumes stationarity, due to this difference the augmented Dickey-Fuller test tends to favour differencing as pointed out in [21]. Applying differencing can cause unwanted dependency in the time series data that did not exist beforehand therefore keeping the model order as low as possible is advisable [61].

As discussed in Subsection 2-2-2, ARMA-type models are typically applied under the assumption that the underlying time series is stationary. This means that the statistical properties of the process, such as mean and variance, are assumed to remain constant over time. If the underlying data do not exhibit these properties, at least locally, the dynamics cannot be effectively captured.

After stationarity is achieved by differencing and/or seasonal differencing, the differenced series are examined using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). This autocorrelation analysis guides the selection of candidate models orders, p, q, p_s, q_s respectively as defined in (2-10), from which several SARIMA model structures are proposed for further evaluation. To balance model accuracy and simplicity, the corrected Akaike information criterion is employed for analysis of the selected model orders. The corrected Akaike information criterion corrects bias in the Akaike information criterion for over estimating the number of model parameters needed and is defined as [59, 62]

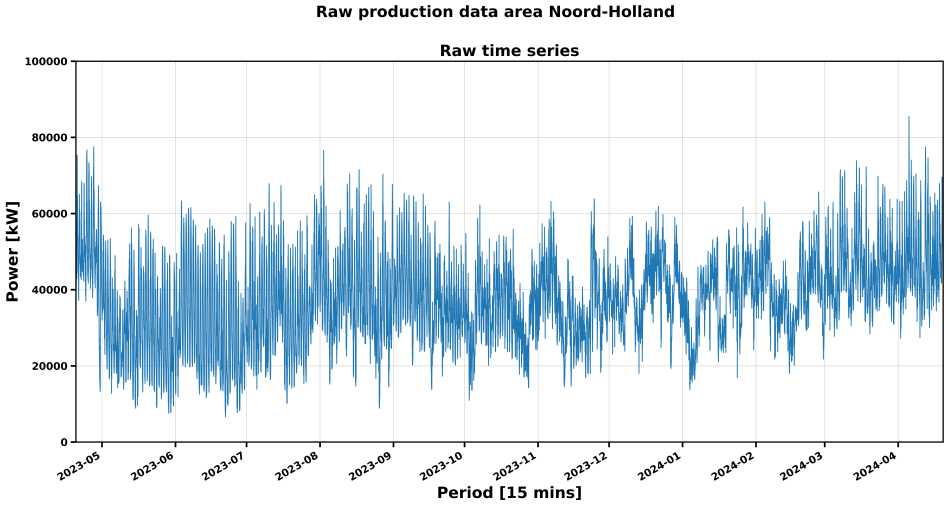
$$\mathrm{AICc} = \mathrm{AIC} + \frac{2(\rho+1)(\rho+2)}{T-\rho-2}, \quad \mathrm{AIC} = -\log(\hat{L}), \tag{3-1}$$

where ρ denotes the total number of estimated parameters in the model, T is the number of usable observations in the time series, and $\hat{L}$ is the maximized likelihood of the fitted model. The additional correction term penalizes models more heavily when the sample size is small relative to the number of parameters, reducing the risk of over-fitting. The model with the lowest corrected Akaike information criterion value is therefore preferred, as it provides the best trade-off between predictive accuracy and model simplicity.

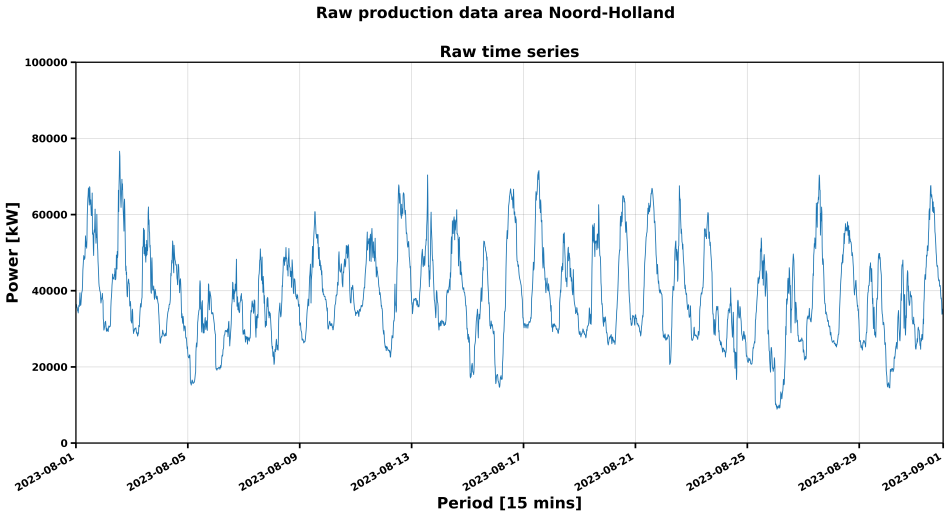
3-2-1 Raw Data Analysis

The data analysis procedure is demonstrated for a single case. Equivalent analyses for the remaining regions are omitted due to space constraints, however the performance of the other fitted models are also shown. The raw production and consumption time series for Noord-Holland are shown in Figures 3-1 and 3-2. Each figure include a full year of data (20th of April 2023 – 20th April 2024) and a zoomed-in view of a August 2023 to better show short-term dynamics.

Figure 3-1 reveals that production exhibits substantial variability driven by renewable generation. These fluctuations occur at multiple periodicities, with frequent rapid spikes superimposed on broader seasonal patterns. Conversely, as seen in Figure 3-2, electricity consumption displays a more structured temporal profile. Daily cycles, weekly working-day effects, and smoother seasonal changes dominate the load pattern, and extreme fluctuations are far less pronounced compared to production. Historically, short-term load forecasting at the



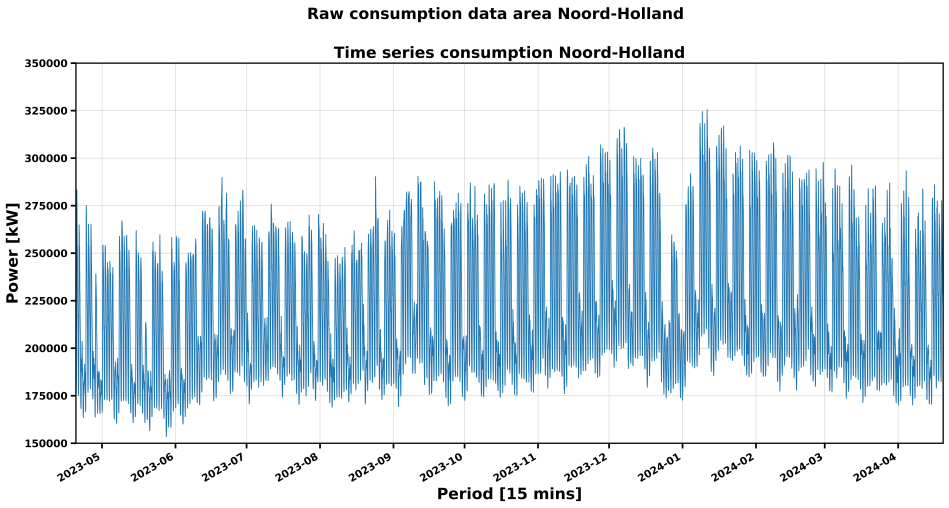
(a) Raw electricity production data for Noord-Holland for April 2023 to April 2024.



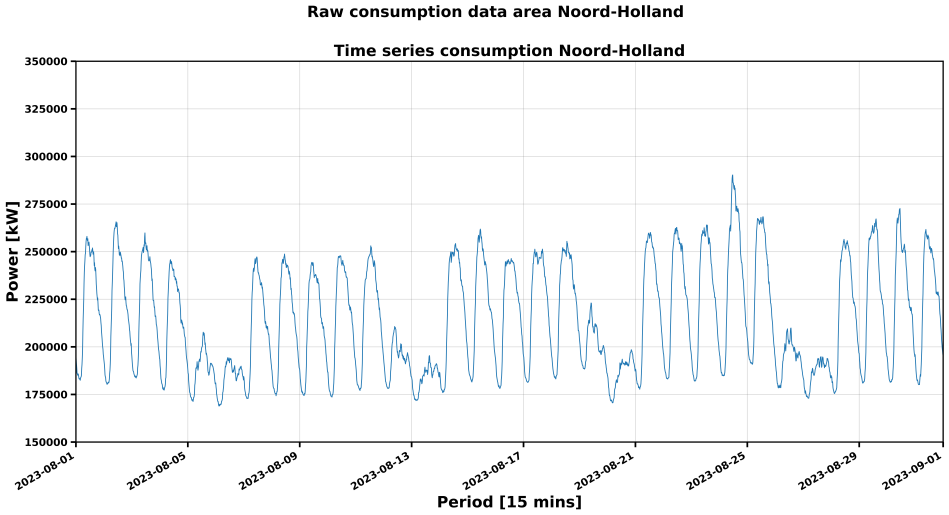
(b) Raw electricity production data for Noord-Holland for August 2023 showing a clear, but highly variable, daily pattern

Figure 3-1: Raw production data for Noord-Holland.

transmission level has been highly accurate due to the predictable nature of aggregated consumption [63]. For example, real-time demand forecasts in Australia achieved 1.88% Mean



(a) Raw electricity consumption data for Noord-Holland for April 2023 to April 2024.



(b) Raw electricity consumption data for Noord-Holland for August 2023, showing clear week and weekend pattern (5 high peaks (week days), followed by 2 low peaks (weekend days)

Figure 3-2: Raw consumption data for Noord-Holland.

Absolute Precentage Error (MAPE) [64], and day-ahead errors as low as 1.36% MAPE were reported in [65]. Given the high accuracy achieved when modelling demand and the clear structural patterns in consumption profiles, this study does not develop separate statistical models for demand. Instead, real measured consumption values are used as input, and the predictive analysis focuses only on the local electricity production from RESs, where volatility results substantial uncertainty.

3-2-2 Stationarity Assessment, Differencing, Seasonality Identification

The raw production data for Noord-Holland are first tested for stationarity using the augmented Dickey–Fuller and Kwiatkowski–Phillips–Schmidt–Shin tests, with the results summarized in Table 3-1. While the augmented Dick-Fuller test suggests that the undifferenced data may already be stationary, the Kwiatkowski–Phillips–Schmidt–Shin test rejects this, indicating the presence of non-stationary behaviour. This non-stationarity conclusion is also consistent with the visibly strong periodic patterns, i.e. changing mean, visible in Figure 3-1. To quantify the dominant recurring behaviour, the seasonal frequency of the data is

	Augmented Dick-Fuller	Kwiatkowski–Phillips–Schmidt–Shin
$d=0,\ D=0$	Stationary	Not stationary
$d=0,\ D=1$	Stationary	Stationary

Table 3-1: Results of stationarity tests applied to the production data of Noord-Holland.

determined using the fast Fourier transform, following the approach in [60]. The spectrum in Figure 3-3 reveals a clear peak corresponding to a cycle of 96 time steps, which is equivalent to a daily seasonality (96 intervals $\times$ 15 minutes = 24 hours). After applying seasonal differencing using (2-11) with the $s=96$ and $d_s=1$, both the ADF and KPSS tests confirm that the transformed series is stationary (Table 3-1).

Based on these findings, the differencing orders are selected as $d=0$ and $D=1$. With stationarity ensured, the next step is to identify suitable model orders, which will be addressed in the following section through analysis of the autocorrelation structures of the preprocessed data.

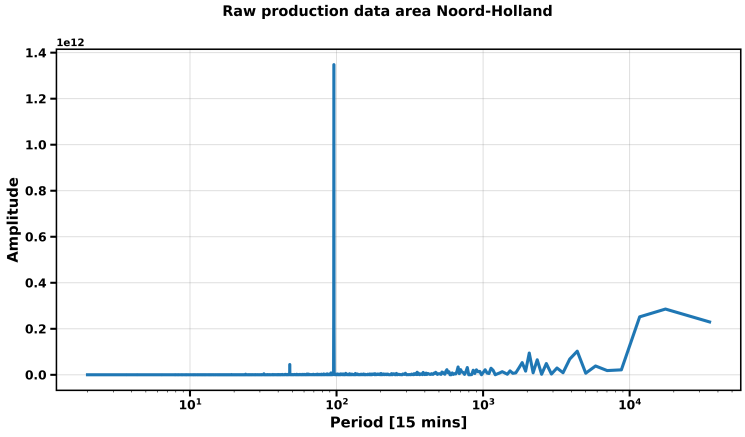


Figure 3-3: FFT spectrum of the production data for Noord-Holland, showing a clear daily periodicity with a dominant frequency corresponding to 96 time steps (24 hours).

3-2-3 Autocorrelation Analysis for Model Order Selection

After applying seasonal differencing, the autocorrelation structure of the differenced production data is examined using the ACF and PACF, shown in Figure 3-4.

Figure 3-4a shows a decaying ACF spikes outside the significant values, where the PACF with a dominant spike at lag 1. This pattern indicates short-memory dependence typical of a low-order autoregressive process, suggesting a small non-seasonal auto-regressive term p component and no clear evidence of a moving-average term $q = 0$.

In Figure 3-4b) the ACF exhibits pronounced negative spikes at 96, while the PACF decays gradually without a clear cut-off. This pattern is characteristic of a seasonal moving-average process of order one, indicating that one seasonal difference ($D = 1$) is sufficient and that the remaining seasonal structure can be captured with a seasonal MA(1) term. Together, these features justify considering models of the general form:

$$\text{SARIMA}(p, 0, 0,\ 0, 1, 1, 96),$$

where the non-seasonal AR(p) component captures short-term dependence and the seasonal MA(1) component accounts for residual autocorrelation at the seasonal frequency.

3-2-4 Model Selection using correct Akaike information criterion

The identified model types are fitted using maximum likelihood estimation and the residuals are evaluated using the correct Akaike information criterion. The lowest correct Akaike information criterion is best therefore the selected model is SARIMA$(5, 0, 0, 0, 1, 1, 96)$.

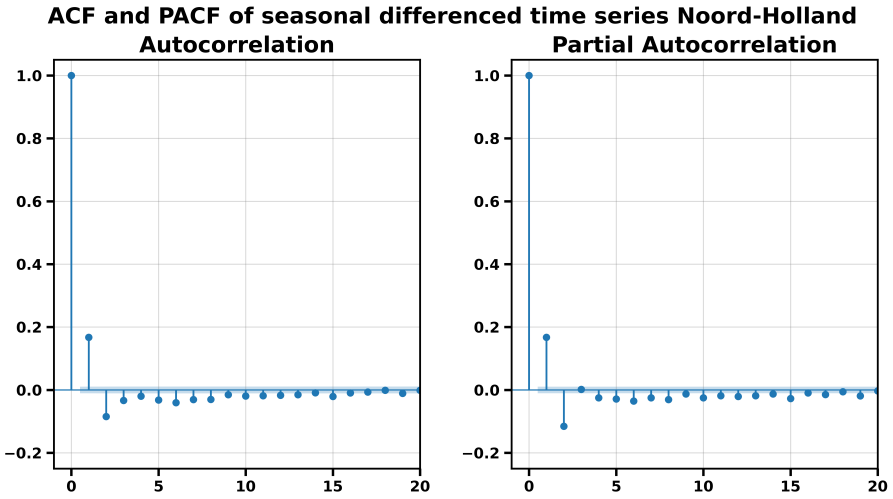
Model	correct Akaike information criterion
SARIMA(1,0,0,0,1,1,96)	618156.84
SARIMA(2,0,0,0,1,1,96)	616888.29
SARIMA(3,0,0,0,1,1,96)	616578.49
SARIMA(4,0,0,0,1,1,96)	616575.48
SARIMA(5,0,0,0,1,1,96)	616569.23

Table 3-2: Results information criterion tests on selected modelling parameters for Noord-Holland

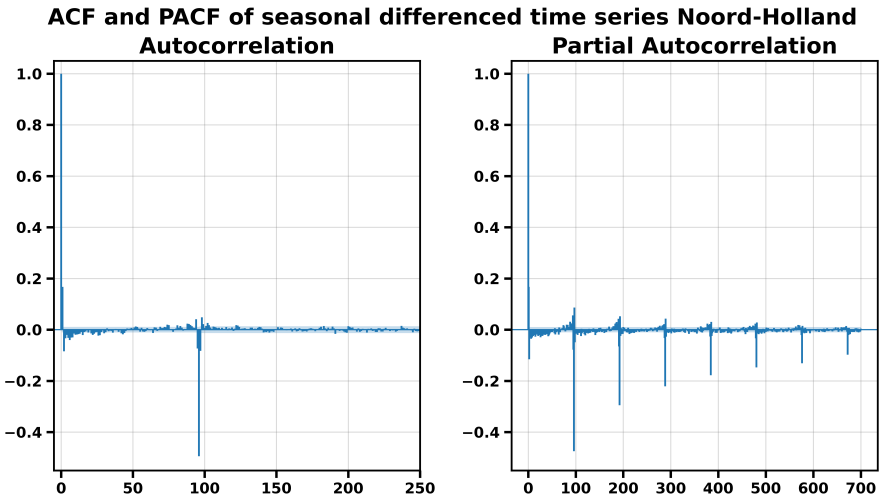
3-3 Model validation

Finally, with the best model parameters selected, the model diagnostics are assessed to determine if the selected SARIMA adequately captures the dynamics. First, the residual time-series plot is examined to visually check for bias and constant variance. The distributional of the residuals are plotted with a histogram over which a Gaussian distribution estimate is plotted, together with a QQ (quantile-quantile) plot, which compares the empirical quantiles of the residuals against a theoretical Gaussian distribution to detect skewness or heavy-tailed behaviour. Finally, the ACF plot is used to test for remaining temporal dependence. These diagnostics collectively verify whether the model is suitable for scenario generation and forecasting.

In addition to residual diagnostics, the predictive accuracy of the SARIMA model is evaluated using an out-of-sample forecasting. The models are used to predict unseen data. The forecast errors are then quantified using the MAPE metric, which measures the average magnitude of the prediction error relative to the observed values.



(a) Zoomed in ACF and PACF



(b) ACF and PACF plot for Noord-Holland

Figure 3-4: Autocorrelation and partial autocorrelation plots of the seasonally differenced time series for Noord-Holland. Subfigure (a) shows a zoomed-in view of the first 20 lags, while subfigure (b) displays the full lag range to illustrate seasonal correlation structures.

Gaussian Noise and Independence Tests

Figure 3-5 presents the residual diagnostics for the fitted SARIMA model on the generation data of Noord-Holland. In the residual time-series plot (top-left), the mean of the residuals fluctuates around zero, indicating no obvious bias. However, noticeable variations in the spread throughout the period suggest heteroscedasticity, meaning the residual variance is not constant over time. This is to be expected due to the seasonal and weather dependency of RES.

The residual distribution is further examined in Figure 3-5. The histogram of the residuals

with an overlaid Gaussian estimate (top-right) shows an approximately bell-shaped form. This observation is supported by the QQ plot (bottom-left), where the theoretical quantiles on the x-axis correspond to a standard normal distribution, while the sample quantiles on the y-axis reflect the actual scale of the residuals. The difference in axis ranges arises because the residuals are not standardized, but the alignment of points along the red reference line still indicates approximate normality. However, slight deviations at the tails suggest the presence of occasional large residuals, implying heavier tails than those expected under a Gaussian distribution.

Serial correlation is examined in the ACF plot (bottom-right), which shows that no significant autocorrelations remain. This indicates that the SARIMA model has effectively captured the underlying patterns in the data, leaving no systematic structure unexplained. However, the residual variance remains high, which is expected to limit the predictive accuracy of the model.

Overall, these diagnostics indicate that the SARIMA model provides unbiased residuals with near Gaussian behaviour and minimal autocorrelation, making it suitable for prediction. At the same time, the observed heteroscedasticity highlights that uncertainty varies with system conditions, thus motivating the use of adaptive variance estimation within the proposed CC-MPC-RL framework, which will be presented in Chapter 5.

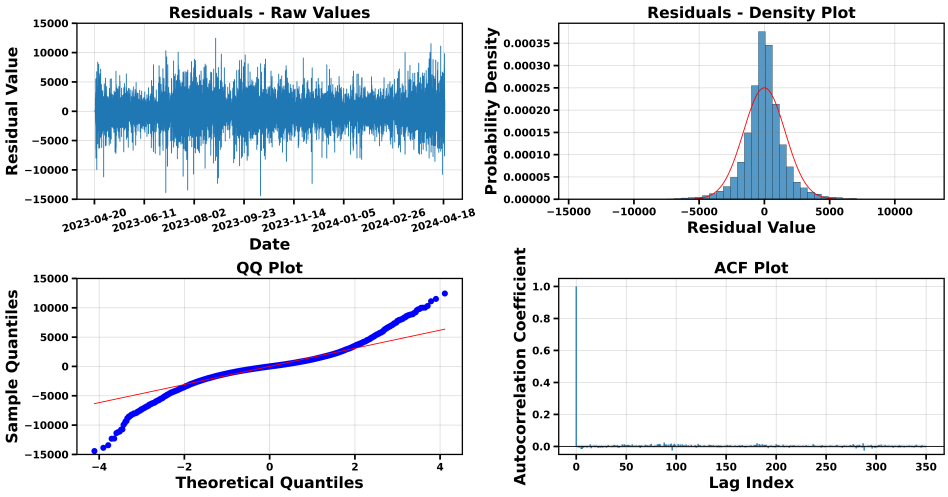


Figure 3-5: Residual diagnostic plots for the fitted SARIMA model: (top-left) residual time series, (top-right) residual histogram with kernel density estimate, (bottom-left) QQ plot, and (bottom-right) .

Out-of-Sample Performance Evaluation

The SARIMA models will be used to make short term generation predictions for the MPC-based controllers that will be derived in Chapter 5. Therefore the predictive accuracy will be tested on the prediction horizons chosen for the methods $n_p \in \{8, 16\}$. After fitting the time-series model, the estimated autoregressive and moving-average parameters are used together

with the distribution of the residuals. First, the most recent values of the seasonally differenced time series and past error terms are stored as initial conditions. At each simulation step, a new forecast value is computed by summing the contributions from the non-seasonal AR and MA terms, as well as the seasonal SAR and SMA components extracted from the fitted model. A random innovation term is then sampled from the fitted normal distribution shown in Figure 3-5. By iterating this process over the desired forecast horizon, one simulated trajectory is generated. The predictions are made by exploiting the recursive structure of the SARIMA model equations, for the final model parameters the model can be formulated as

$$y_{t+1} = \sum_{i=1}^{p} \phi_i \, y_{t+1-i} + \sum_{j=1}^{q_s} \Theta_j \, \varepsilon_{t+1-j} + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathcal{N}(0, \sigma^2), \tag{3-2}$$

where ϕ_i and Θ_j denote the non-seasonal AR and seasonal MA coefficients, s the seasonal period, and ε_{t+1} a random innovation drawn from the fitted Gaussian distribution. By iterating this recursion over the desired horizon, one simulated trajectory is obtained. Repeating the process with newly sampled innovations yields multiple different paths. The forecasting

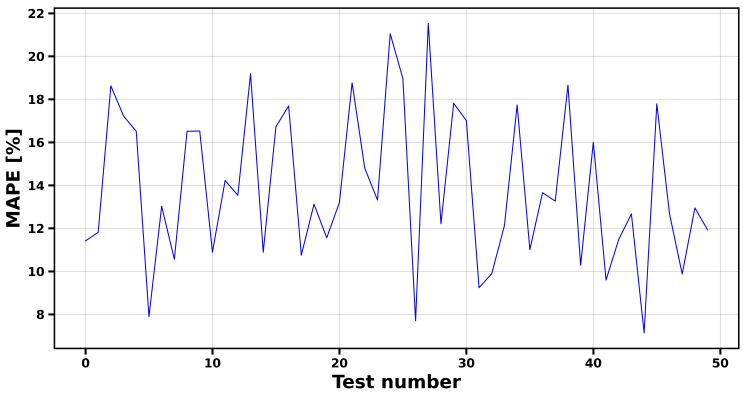


Figure 3-6: MAPE of ARMA across 50 simulated days

results for all regions are summarised in Table 3-3. The forecasting results for all regions

Province	RMSE (mean)	MAPE	MAPE (variance)
Noord-Holland	6497 $[kW]$	13.86 %	8.64 %
Zuid-Holland	26608 $[kW]$	23.25 %	20.02 %
Groningen	22882$[kW]$	495.45 %	4959.55 %
Zeeland	11765 $[kW]$	31.69 %	32.00%
Brabant	17001 $[kW]$	19.89 %	18.80 %
Limburg	9965 $[kW]$	23.03 %	19.24 %
Utrecht, Flevopolder, en Gelderland	19864 $[kW]$	27.68 %	27.82 %
Friesland	7679 $[kW]$	41.94 %	53.05 %

Table 3-3: Out-of-sample prediction accuracy per region for prediction horizon $n_p = 16$

are summarised in Table 3-3. The Root Mean Square Error (RMSE) values indicate that the absolute forecast errors remain within a reasonable range for short-term renewable generation

prediction. Most regions show MAPE values between 10 % and 30 %, which unfortunately is not very accurate but to expected due too the high variability of the data since all different kinds of RES and consumption are bundled together.

Two outliers appear in the MAPE scores: Groningen and Friesland. They exhibit significantly higher MAPE values. In Groningen, this is primarily due to periods of very low or zero production, which causes percentage-based errors to inflate despite relatively normal absolute deviations. Friesland's higher MAPE Also has periods of lower production periods which drive percentage based errors up but no near zero values to explode the number as big as for Groningen.

Despite the elevated MAPE values observed in these regions, the corresponding RMSE values remain within an admissible range on the total scale of the network. Therefore, the forecasting models remain usable within the proposed CC-MPC framework.

3-4 Summary

This chapter presents the data foundation and uncertainty modelling used in the proposed CC-MPC framework. Aggregated electricity production and consumption data were obtained for eight Dutch provinces. Due to the high predictability of aggregated demand, real consumption values are used directly, while forecasting efforts focus on the more volatile renewable production component.

A systematic time-series analysis workflow is applied to develop a stochastic generation forecast model. Raw data characteristics are examined to identify trends, seasonality, and variability. Stationarity is assessed using augmented Dick–Fuller and Kwiatkowski–Phillips–Schmidt–Shin tests, and confirmed after applying daily seasonal differencing, guided by dominant frequencies identified via fast Fourier transform. The autocorrelation structure of the stationary data is then analysed using ACF and PACF plots to propose candidate SARIMA model orders.

Model selection is performed using the corrected Akaike information criterion to balance forecasting accuracy and model complexity. The selected model undergoes residual analysis, confirming near-Gaussian white-noise behaviour and appropriate capture of temporal structure. Out-of-sample forecasting shows the out-of-sample predictive accuracy of the models. Overall, the chapter develops a statistical forecasting model for RESs production, enabling its use within the CC-MPC-RL control framework introduced later in the thesis.

Chapter 4

Modelling

In this chapter, the foundation for the MPC implementation is established. Section 4-1 presents the linearised dynamical model of the high-voltage transmission network, which serves as the system representation within the control framework. Section 4-2 then derives the market model, based on the Dutch congestion market GOPACS, that integrates with the proposed methodology.

4-1 Linearised dynamical transmission network model

As discussed in Chapter 2, the physical model of the transmission network is non-linear. To enable its use within the MPC framework, a linearised time-varying representation is derived here. The resulting model takes the following form:

$$x(k+1) = A(k)x(k) + B(k)u(k) + B(k)w(k), \tag{4-1}$$

where $x(k)$ is the current real and reactive power output of each node and real and reactive power transmissions for each transmission line. The control inputs $u(k)$ are the net changes of power at each node due to the activation of market offers and $w(k)$ are all the other changes in power at each node as these are not controllable from the market. The state, control and disturbance vectors are defined as follows:

$$\begin{aligned} x(k) = \big[&P^{(1)}(k) \dots P^{(|\mathscr{N}|)}(k) \;\; Q^{(1)}(k) \dots Q^{(|\mathscr{N}|)}(k) \\ &P^{(1,2)}(k) \dots P^{(|\mathscr{E}|)}(k) \;\; Q^{(1,2)}(k) \dots Q^{(|\mathscr{E}|)}(k)\big]^T \end{aligned} \tag{4-2a}$$

$$u(k) = \big[\Delta P_u^{(1)}(k) \dots \Delta P_u^{(|\mathscr{N}|)}(k) \;\; \Delta Q_u^{(1)}(k) \dots \Delta Q_u^{(|\mathscr{N}|)}(k)\big]^T \tag{4-2b}$$

$$w(k) = \big[\Delta P_w^{(1)}(k) \dots \Delta P_w^{(|\mathscr{N}|)}(k) \;\; \Delta Q_w^{(1)}(k) \dots \Delta Q_w^{(|\mathscr{N}|)}(k)\big]^T, \tag{4-2c}$$

where the states consist of the net real $P^{(n)}(k)$ and reactive $Q^{(n)}(k)$ power generation or demand of each node $n \in \mathscr{N}$, and the net real $P^{(n,m)}(k)$ and reactive power flow $Q^{(n,m)}(k)$ of each transmission line $(n,m) \in \mathscr{E}$. The control inputs are the controllable real $\Delta P_u^{(n)}(k)$ and

reactive power changes $\Delta Q_u^{(n)}(k)$ at node n. The disturbance are all other changes in real $\Delta P_w^{(n)}(k)$ and reactive power $\Delta Q_w^{(n)}(k)$. The dynamics of real and reactive power at a node is defined as

$$P^{(n)}(k+1) = P^{(n)}(k) + \Delta P_u^{(n)}(k) + \Delta P_w^{(n)}(k) \quad \forall n \in \mathscr{N} \tag{4-3a}$$

$$Q^{(n)}(k+1) = Q^{(n)}(k) + \Delta Q_u^{(n)}(k) + \Delta Q_w^{(n)}(k) \quad \forall n \in \mathscr{N}. \tag{4-3b}$$

The dynamics of the power transmission are derived from the non-linear relation in (2-3), which is approximated using a first order Taylor series expansion. Due to space constraints the full derivation is omitted here and the reader is referred to Appendix A. The linearised model for change of power transmission is defined as follows [66]

$$\Delta P^{(n,m)}(k) = \sum_{l \in \mathscr{N}} g_{\mathrm{pp}}^{(n,m),l}(k) \Delta P^l(k) + g_{\mathrm{pq}}^{(n,m),l}(k) \Delta Q^l(k) \quad \forall (n,m) \in \mathscr{E} \tag{4-4a}$$

$$\Delta Q^{(n,m)}(k) = \sum_{l \in \mathscr{N}} g_{\mathrm{qp}}^{(n,m),l}(k) \Delta P^l(k) + g_{\mathrm{qq}}^{(n,m),l}(k) \Delta Q^l(k) \quad \forall (n,m) \in \mathscr{E} \tag{4-4b}$$

where $g_{\mathrm{pp}}^{(n,m),l}(k)$ are AC power transfer distribution factors from real power change in node l to real power transmission in line (n,m) similar to the work in [66]. Combining (4-3) and (4-4) into one model results in the following model in the form of (4-1)

$$\begin{aligned}
\begin{bmatrix} \boldsymbol{P}^{(n)}(k+1) \\ \boldsymbol{Q}^{(n)}(k+1) \\ \boldsymbol{P}^{(n,m)}(k+1) \\ \boldsymbol{Q}^{(n,m)}(k+1) \end{bmatrix} = & \underbrace{I_{4|\mathscr{N}|\times 4|\mathscr{N}|}}_{A(k)} \begin{bmatrix} \boldsymbol{P}^{(n)}(k) \\ \boldsymbol{Q}^{(n)}(k) \\ \boldsymbol{P}^{(n,m)}(k) \\ \boldsymbol{Q}^{(n,m)}(k) \end{bmatrix} + \underbrace{\begin{bmatrix} I_{|\mathscr{N}|\times|\mathscr{N}|} & 0 \\ 0 & I_{|\mathscr{N}|\times|\mathscr{N}|} \\ G(k)^{(PP)} & G(k)^{(PQ)} \\ G(k)^{(QP)} & G(k)^{(QQ)} \end{bmatrix}}_{B(k)} \begin{bmatrix} \Delta \boldsymbol{P}_u^{(n)}(k) \\ \Delta \boldsymbol{Q}_u^{(n)}(k) \end{bmatrix} \\
& + \underbrace{\begin{bmatrix} I_{|\mathscr{N}|\times|\mathscr{N}|} & 0 \\ 0 & I_{|\mathscr{N}|\times|\mathscr{N}|} \\ G_{\mathrm{pp}}(k) & G_{\mathrm{pq}}(k) \\ G_{\mathrm{qp}}(k) & G_{\mathrm{qq}}(k) \end{bmatrix}}_{B(k)} \begin{bmatrix} \Delta \boldsymbol{P}_w^{(n)}(k) \\ \Delta \boldsymbol{Q}_w^{(n)}(k) \end{bmatrix}
\end{aligned} \tag{4-5}$$

where the vectors $\boldsymbol{P}^{(n)}(k), \boldsymbol{Q}^{(n)}(k), \boldsymbol{P}^{(n,m)}(k), \boldsymbol{Q}^{(n,m)}(k)$ are the collection of all individual variables and matrices $G_{\mathrm{pp}}(k)$, $G_{\mathrm{pq}}(k)$, $G_{\mathrm{qp}}(k)$, and $G_{\mathrm{qq}}(k)$ are defined as in (A-23). The model in (4-5) is for a single time step but can easily be adapted to include all time steps in the prediction horizon in the following way

$$\mathbf{x}(k+1) = S(k)x(k) + T(k)\boldsymbol{u}(k) + T(k)\boldsymbol{w}(k) \tag{4-6}$$

where the matrices $S(k)$ and $T(k)$ are defined as

$$S(k) = \begin{bmatrix} I_{4|\mathscr{N}|\times 4|\mathscr{N}|} \\ \vdots \\ I_{4|\mathscr{N}|\times 4|\mathscr{N}|} \end{bmatrix}, T(k) = \begin{bmatrix} B(k) & 0 & 0 \\ \vdots & \ddots & 0 \\ B(k) & \dots & B(k) \end{bmatrix} \tag{4-7}$$

and

$$\boldsymbol{x}(k+1) = \begin{bmatrix} x(k+1) \\ x(k+2) \\ \vdots \\ x(k+n_{\mathrm{p}}) \end{bmatrix}, \boldsymbol{u}(k) = \begin{bmatrix} u(k) \\ u(k+1) \\ \vdots \\ u(k+n_{\mathrm{p}}-1) \end{bmatrix}, \boldsymbol{w}(k) = \begin{bmatrix} w(k) \\ w(k+1) \\ \vdots \\ w(k+n_{\mathrm{p}}-1) \end{bmatrix}, \tag{4-8}$$

where the bold symbols denote the fact that it is the vector over the whole prediction horizon. For the linearisation too hold the voltage angle difference between connected nodes must be small, i.e. $\pm 3.5°$. For high voltage networks this is a valid assumption [29].

4-1-1 Limits

Transmission lines are subject to physical and thermal constraints that restrict the maximum amount of power that can be safely transmitted. Excessive power flow causes line heating, which increases conductor resistance and may lead to permanent damage or tripping of protection systems. To ensure secure operation, these thermal constraints are represented as limits on the apparent power flow in each transmission line.

The apparent power limit, commonly referred to as the thermal limit, defines the maximum apparent power transfer. It is formulated as a quadratic constraint on the apparent power magnitude as follows [67, 68]:

$$\left(P^{(n,m)}(k+i)\right)^2 + \left(Q^{(n,m)}(k+i)\right)^2 \leq \left(S^{\max}\right)^2 \quad \forall i \in \{1, \ldots, n_{\mathrm{p}}\}, (n,m) \in \mathscr{E}, \tag{4-9}$$

where $S^{\max}$ denotes the maximum apparent power transferable through the transmission lines, and $P^{(n,m)}(k+i)$ and $Q^{(n,m)}(k+i)$ represent the corresponding real and reactive power flows at time step $k+i$. These limits are incorporated into the MPC framework to ensure that the optimisation respects the physical operating boundaries of the network.

4-2 Market model

As discussed in Subsection 2-1-2, the Dutch congestion market operates as a pool-based market, where participants submit offers according to the specifications summarized in Table 2-1. In this section, the restrictions imposed on the control inputs by the Dutch market design are formulated within an MPC framework.

The complete set of offers, denoted by $\mathcal{O}$, contains tuples representing the full definition of each offer, which will be detailed in the following subsections. A single offer is denoted by PO^{o} or $\mathrm{FTO}^{\mathrm{o}}$, depending on its type, where o is the offer index and $\mathrm{PO}^{\mathrm{o}}, \mathrm{FTO}^{\mathrm{o}} \in \mathcal{O}$. The set $\mathcal{O}$ can be divided into two subsets based on the offer type: the Profile Offers (POs), i.e., offers specifying fixed power profiles over time, and the Flex-Time Offers (FTOs), i.e., offers that allow temporal flexibility in the activation of power. These subsets are defined as follows:

$$\mathcal{O}^{\mathrm{PO}} \subseteq \mathcal{O}, \quad \mathcal{O}^{\mathrm{FTO}} \subseteq \mathcal{O}, \quad \mathcal{O}^{\mathrm{PO}} \cap \mathcal{O}^{\mathrm{FTO}} = \emptyset. \tag{4-10}$$

Each offer $o \in \mathcal{O}$ belongs exclusively to one of these two subsets, ensuring that POs and FTOs are mutually exclusive and together form the complete set of offers.

4-2-1 Balancing requirement

The overall control input can be decomposed into the contributions from each individual offer. Hence, the total control action at time step k is given by

$$u(k+i) = \sum_{\mathrm{o} \in \mathcal{O}} u^{(\mathrm{o})}(k+i) \quad \forall i \in \{0, \ldots, n_{\mathrm{p}} - 1\}, \tag{4-11}$$

where $u^{(o)}(k+i)$ is the control action associated to the offer indexed by o, and it is zero for every node that is not the node at which the offer is made, as shown in

$$u^{(o)}(k+i) = \begin{cases} [0 \dots 0 \ \Delta P_u^{(n)}(k+i) \ 0 \dots 0]^T & \forall k+i \in \{t^{\text{start},(o)}, \dots, t^{\text{stop},(o)}\} \\ [0 \dots 0]^T & \text{otherwise,} \end{cases} \tag{4-12}$$

where the change of power $\Delta P_u^{(n)}(k+i), t^{\text{start},(o)}$, and $t^{\text{stop},(o)}$ associated with each offer type o will be defined in Subsections 4-2-2 and 4-2-3. Since the congestion market operates independently of the balancing market, the net effect of congestion management actions must not alter the system's power balance. This constraint can be expressed as

$$\sum_{o \in \mathcal{O}} u^{(o)}(k+i) = 0 \quad \forall i \in \{0, \dots, n_p - 1\}, \tag{4-13}$$

ensuring that the aggregate control actions across all offers remain power neutral at all times. As the congestion market in the Netherlands only considers real power offers these are the only one under consideration even thought the model also allows for reactive power offers.

4-2-2 Profile offers

RESs are weather-dependent and therefore cannot shift their flexibility in time. To model curtailment or other similar one-time flexibility actions, the corresponding offers must be valid only for a single moment and must specify a fixed power profile for each time step. Such offers are represented by the tuple

$$\text{PO}^{(o)} : (n, t^{\text{start},(o)}, t^{\text{stop},(o)}, \beta^{\min,(o)}, P^{(o)}, c^{(o)}, m^{(o)}) \tag{4-14}$$

where n denotes the node index associated with offer o. The start and end time of the offer are denoted by $t^{\text{start},(o)}$ and $t^{\text{stop},(o)}$ respectively. The minimum activation fraction is defined as $\beta^{\min,(o)}$. The power profile associated with offer o, denoted by $P^{(o)}$, defines the amount flexible power at each time step within the offer's start and end time. Finally, the cost is denoted by $c^{(o)}$ and market direction by $m^{(o)} \in \{-1, 1\}$ with $+1$ for a buy bid and -1 for a sell bid. The power profile $P^{(o)}$ is defined as

$$P^{(o)} = [P_u^{(n)}(t^{\text{start},(o)}) \dots P_u^{(n)}(t^{\text{stop},(o)})] \tag{4-15}$$

where $P_u^{(n)}(t^{\text{start},(o)})$ represents the scheduled power output at node n and time step $t^{\text{start},(o)}$. The change in power output resulting from activating a given offer is defined as

$$\Delta P_u^{(n),(o)}(k+i) = \begin{cases} (P_u^{(n),(o)}(k+i+1) - P_u^{(n),(o)}(k+i))\beta^{(o)}\delta^{(o)} & \forall k+i \in \{t^{\text{start}}, \dots, t^{\text{stop}}+1\} \\ 0 & \text{otherwise.} \end{cases} \tag{4-16}$$

Here, $\beta^{(o)} \in [\beta^{\min,(o)}, 1]$ is a continuous scaling factor determining the magnitude of activation, whereas $\delta^{(o)}$ is a single binary variable indicating whether the offer is activated. The product $\beta^{(o)}\delta^{(o)}$ introduces a non-linear relationship. To address the non-linearity introduced in (4-16), a reformulation is applied to linearise the corresponding terms in the optimisation problem,

following the standard approach described in [69].

$$\begin{aligned}
\Delta P_u^{(n),(\mathrm{o})}(k+i) &\leq M^{\mathrm{high}}\delta^{(\mathrm{o})} - \sum_{j=t_{\mathrm{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(j) \\
\Delta P_u^{(n),(\mathrm{o})}(k+i) &\geq M^{\mathrm{low}}\delta^{(\mathrm{o})} - \sum_{j=t_{\mathrm{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(k+j) \\
\Delta P_u^{(n),(\mathrm{o})}(k+i) &\leq \beta^{(\mathrm{o})} P^{(n),(\mathrm{o})}(k+i) - \sum_{j=t_{\mathrm{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(k+j) - M^{\mathrm{low}}(1-\delta^{(\mathrm{o})}) \\
\Delta P_u^{(n),(\mathrm{o})}(k+i) &\geq \beta^{(\mathrm{o})} P^{(n),(\mathrm{o})}(k+i) - \sum_{j=t_{\mathrm{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(k+j) - M^{\mathrm{high}}(1-\delta^{(\mathrm{o})})
\end{aligned} \tag{4-17}$$

In this formulation, M^{high} and M^{low} represent sufficiently large positive and negative constants that define the upper and lower bounds of the feasible range for $\Delta P_u^{(n)}(k)$. When the offer is inactive ($\delta^{(\mathrm{o})} = 0$), the inequalities force $\Delta P_u^{(n)}(k)$ to zero. Conversely, when $\delta^{(\mathrm{o})} = 1$, the equations allow $\Delta P_u^{(n)}(k)$ to take on values consistent with the power change determined by $\beta^{(\mathrm{o})}$.

4-2-3 Flex-time offers

Flexible assets such as batteries are often characterized by a maximum power, which can be delivered when necessary for as long as the battery is not empty. To capture this behaviour within the market model, the FTO is introduced. An FTO represents an offer with a constant maximum power but a flexible activation period, allowing the market mechanism to optimally allocate its operation in time. This offer is defined by the tuple

$$\mathrm{FTO}^{(\mathrm{o})} : (n, t^{\mathrm{start},(\mathrm{o})}, t^{\mathrm{stop},(\mathrm{o})}, \ell^{\mathrm{min},(\mathrm{o})}, \ell^{\mathrm{max},(\mathrm{o})}, \beta^{\mathrm{min},(\mathrm{o})}, P^{\mathrm{max},(\mathrm{o})}, c^{(\mathrm{o})}, m^{(\mathrm{o})}) \tag{4-18}$$

where n denotes the node index associated with offer o. The set of all FTO offers is denoted by $\mathcal{O}^{\mathrm{FTO}}$ and each element $\mathrm{FTO}^{\mathrm{o}} \in \mathcal{O}^{\mathrm{FTO}}$ represents a single FTO. The parameters $t^{\mathrm{start},(\mathrm{o})}$ and $t^{\mathrm{stop},(\mathrm{o})}$ define the earliest activation and latest possible end time, while $\ell^{\mathrm{min},(\mathrm{o})}$ and $\ell^{\mathrm{max},(\mathrm{o})}$ specify the minimum and maximum consecutive activation periods. The variables $\beta^{\mathrm{min},(\mathrm{o})}$ and $P^{\mathrm{max},(\mathrm{o})}$ denote the minimum activation fraction and maximum power quantity. The offer cost is denoted by $c^{\mathrm{offer},(\mathrm{o})}$, and $m^{\mathrm{offer},(\mathrm{o})}$ indicates the market direction, with $+1$ for buy and -1 for sell offers. For a single offer the constraint on the change of power in a node is then defined as

$$\Delta P_u^{(n)}(k+i) = \begin{cases} \left(\delta^{(\mathrm{o})}(k+i) - \delta^{(\mathrm{o})}(k+i-1)\right)\beta^{(\mathrm{o})} P^{\mathrm{max},(\mathrm{o})} & \forall k+i \in \{t^{\mathrm{start},(\mathrm{o})}, \dots t^{\mathrm{stop},(\mathrm{o})}\} \\ 0 & \text{otherwise,} \end{cases} \tag{4-19}$$

where $\Delta P_u^{(n)}(k)$ represents the change in power at node n in time step k. The binary variable $\delta^{(\mathrm{o})}(k)$ indicates the activation status of the offer at time step k, taking the value 1 when the offer is active at time step k and 0 otherwise. The parameter $\beta^{(\mathrm{o})} \in [\beta^{\mathrm{min},(\mathrm{o})}, 1]$ defines the activation fraction of the offer, with $+1$ for a buy bid and -1 for a sell bid, while $P^{\mathrm{max},(\mathrm{o})}$ denotes the corresponding offered power associated with $\mathrm{FTO}^{(\mathrm{o})}$.

Once an FTO is activated, it should remains active for at least the minimum duration $\ell_{\min}$. This condition is enforced through the binary activation variable $\delta^{(\mathrm{o})}(k)$, which indicates whether offer (o) is active at time step k. This behaviour is encoded in the following constraint

$$\sum_{j=i}^{k+\ell_{\min}-1} \delta^{(\mathrm{o})}(k+j) \geq \ell_{\min}\big(\delta^{(\mathrm{o})}(k+i) - \delta^{(\mathrm{o})}(k+i-1)\big) \quad \forall k+i \in \{t_{\text{start}}, \dots, t_{\text{stop}} - \ell_{\min} + 1\}, \tag{4-20}$$

which ensures that if the activation variable switches from 0 at time $k+i-1$ to 1 at time $k+i$, the offer remains active for at least $\ell_{\min}$ consecutive time steps. In other words, each activation must last no shorter than the minimum duration defined in the FTO. To enforce the maximum activation duration, a limit $\ell_{\max}$ is imposed to prevent an offer from remaining active beyond the allowable time. This condition is expressed a

$$\sum_{j=t_{\text{start}}}^{t_{\text{stop}}} \delta_j^{(\mathrm{o})} \leq \ell^{\max}, \tag{4-21}$$

which restricts the total number of time periods during which the offer can be active to the specified limit. However, this formulation still permits multiple non-consecutive activations within the activation window. To prevent such cases, the following constraint is added:

$$\sum_{j=i}^{t^{\text{stop}}} \delta_j^{(\mathrm{o})} \leq \ell^{\max}\big(1 + \delta^{(\mathrm{o})}(k+i) - \delta^{(\mathrm{o})}(k+i-1)\big) \quad \forall k+i \in \{t^{\text{start}} + \ell^{\min}, \dots, t^{\text{start}} + \ell^{\max} - 1\}, \tag{4-22}$$

which ensures that once the activation variable switches from 1 at time $k+i-1$ to 0 at time $k+i$, all subsequent activation variables remain 0. In other words, this constraint prevents multiple non-contiguous activations of the same offer within the allowed time window. The combination of (4-20), (4-21), and (4-22) make sure that at an FTO is activated at most once with a length between $\ell^{\min}$ and $\ell^{\max}$. The formulation in (4-19) is non-linear due to the multiplication of β and δ. To remove this nonlinearity from the optimisation problem, the constraint is reformulated in the following way [69]

$$\begin{aligned}
\Delta P_u^{(n),(\mathrm{o})}(k+i) &\leq M^{\text{high}}\delta^{(\mathrm{o})}(k+i) - \sum_{j=t^{\text{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(j) \\
\Delta P_u^{(n),(\mathrm{o})}(k+i) &\geq M^{\text{low}}\delta^{(\mathrm{o})}(k+i) - \sum_{j=t^{\text{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(j) \\
\Delta P_u^{(n),(\mathrm{o})}(k+i) &\leq \beta^{(\mathrm{o})} P^{\max,(\mathrm{o})} - \sum_{j=t^{\text{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(j) - M^{\text{low}}(1 - \delta^{(\mathrm{o})}(k+i)) \\
\Delta P_u^{(n),(\mathrm{o})}(k+i) &\geq \beta^{(\mathrm{o})} P^{\max,(\mathrm{o})} - \sum_{j=t^{\text{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(j) - M^{\text{high}}(1 - \delta^{(\mathrm{o})}(k+i)),
\end{aligned} \tag{4-23}$$

where M^{high} and M^{low} are large positive and negative constants, respectively, chosen to bound the feasible range of $\Delta P^{(n,u)}(k+i)$. This formulation ensures that the power adjustment $\Delta P^{(n,u)}(k+i)$ only takes meaningful values when the offer is active (i.e., $\delta^{(\mathrm{o})}(k+i) = 1$), while it is forced to zero when the offer is inactive ($\delta^{(\mathrm{o})}(k+i) = 0$). As a result, the non-linear dependency between $\beta^{(\mathrm{o})}$ and $\delta^{(\mathrm{o})}$ is handled in a linear manner.

4-2-4 Cost

The total volume associated with a given action is determined by the activation fraction $\beta^{\min,(o)}$. Consequently, the cost of an offer corresponds to the product of the total activated volume and the respective offer price. Since the control actions represent changes in power, the total energy exchanged is obtained by summing over all cumulative power adjustments. Furthermore, as the network operator compensates for the difference between the buy and sell offers, the cost is weighted by the offer direction $m^{(\mathrm{o})} \in \{-1, 1\}$, indicating whether the offer represents an increase or decrease in power. The total cost incurred by the network operator is therefore expressed as

$$C^{\mathrm{total},(\mathrm{o})} = \sum_{\mathrm{o}\in\mathcal{O}} m^{(\mathrm{o})} \beta^{(\mathrm{o})} c^{(\mathrm{o})} \sum_{i=1}^{n_{\mathrm{p}}} \sum_{j=1}^{i} u_{k+j}^{(\mathrm{o})} \tag{4-24}$$

where $C^{\mathrm{total},(\mathrm{o})}$ denotes the total cost for the network operator, $m^{(\mathrm{o})}$ is the offer direction, $c^{(\mathrm{o})}$ is the offer price, and $u_{k+j}^{(\mathrm{o})}$ represents the power change associated with offer o at time step j. The inner summation accumulates the power deviations over the prediction horizon, while the outer summation aggregates the corresponding energy costs across all offers in $\mathcal{O}$.

4-3 Summary

In this chapter, the core mathematical models underpinning the thesis have been presented: a linearised dynamical model for a high-voltage transmission network and a market model formulated within an MPC framework. The dynamical model captures real and reactive power dynamics at each node and along transmission lines, derived through a first-order linearisation of the AC power flow equations. This formulation enables the prediction of system behaviour as a foundation for MPC-based CM.

Then, a novel MPC-compatible market formulation inspired by the Dutch GOPACS congestion market is derived. Two offer types, POs and FTOs, are modelled according to the specifications of the high-voltage grid operator. The model uses binary variables to encode the market orders as control actions in the MPC. Together, these models establish the basis for a dynamic and market-based model that enables real-time, MPC-based CM within the existing Dutch congestion market structures.

Chapter 5

Control methods

In this chapter, various control methods for CM are presented. The discussion begins with approaches inspired by existing research, followed by the introduction of the novel control framework developed in this thesis. Section 5-1 describes a greedy matching method for congestion control. Section 5-2 then introduces an market-based MPC-based approach, which is subsequently extended to a CC-MPC formulation in Section 5-3. Finally, Section 5-4 presents the proposed method, which combines the CC-MPC with RL agent to improve performance.

5-1 Algorithmic Greedy Matching approach for CM

The first approach to consider is a greedy matching strategy that removes predicted congestion by activating the cheapest pair of market offers that solves the issue, serving as a practical benchmark for the optimization-based controllers in the next sections. The approach does not capture potential combinations beyond pairwise trades, though it reflects a realistic dispatcher tactic: trying a few cheap counter-trades and immediately validating them against the actual network. Albeit simple, this algorithmic greedy matching approach highlights the gains provided by more advanced MPC and CC-MPC controllers over a more realistic implementation.

At time k, the method uses the current state $x(k)$, the previously scheduled inputs $\boldsymbol{u}(k|k-i)$, where the notation $k|k-i$ indicates that the input was scheduled at time $k-i$, and the predicted disturbances $\boldsymbol{w}(k)$ to perform a power-flow simulation over the next n_{p} time steps. At each step the transmission line power flow magnitudes are checked. If no violations occur no action is needed. If congestion is predicted, each offer pair is checked for the following conditions:

- Directions must be opposite. Two buys or two sells cannot net to zero.
- Offers must lie fully within $[k,\, k+n_{\mathrm{p}}]$.
- Offers must be at different nodes.

For a feasible pair, the power volumes P^{buy} and P^{sell} over their activation window are balanced using

$$\beta^{\text{buy}} = \begin{cases} 1 & \text{if } |P^{\text{buy}}| \leq |P^{\text{sell}}| \\ \frac{|P^{\text{sell}}|}{|P^{\text{buy}}|} & \text{Otherwise} \end{cases}, \beta^{\text{sell}} = \begin{cases} 1 & \text{if } |P^{\text{sell}}| \leq |P^{\text{buy}}| \\ \frac{|P^{\text{buy}}|}{|P^{\text{sell}}|} & \text{Otherwise} \end{cases}, \tag{5-1}$$

so that the maximum matchable power is activated. The total input $\boldsymbol{u}(k|k) = \sum_{i=0}^{n_{\text{p}}} \boldsymbol{u(k|k-i)}$ is then checked. If all previously predicted violations disappear and no new ones arise, the pair is accepted and its activation cost is

$$C^{\text{total}} = \beta^{\text{buy,(o)}} C^{\text{buy}} - \beta^{\text{sell,(o)}} C^{\text{sell}},$$

where C^{buy} and C^{sell} are defined as in (4-24). The total cost is thus defined by the spread between the buy and the sell order. Among all accepted pairs, the minimum cost solution is chosen.

5-2 CM as an MPC problem

Starting from the MPC framework defined in (2-12), the complete formulation can be specified by substituting the corresponding model components as follows. For the cost function (2-12a), the cost expression in (4-24) is applied. The cost function is formulated as an economic MPC where only the financial cost are under consideration [70].

In conventional MPC, only the first control action of the optimized sequence is applied at each time step, and the optimization problem is re-solved at the next step using updated system information. However, in this case, once an offer is activated, it must remain active for its entire duration. Therefore, previously activated control actions must continue to be considered in subsequent optimization steps. The full applied control action $\boldsymbol{u}(k)$ at time k is the sum of the current control action and the shifted control actions from the previous n_{p} time steps, defined as

$$\boldsymbol{u}(k) = \sum_{i=0}^{n_{\text{p}}} \boldsymbol{u}(k|k-i), \tag{5-2}$$

where only $\boldsymbol{u}(k|k)$ is an optimization variable, and the remaining $\boldsymbol{u}(k|k-i)$ terms are treated as constants. To ensure that control actions are correctly aligned in time, the previously determined control sequences are shifted forward at each time step as follows:

$$\boldsymbol{u}(k|k-i) = \begin{bmatrix} u(k+i|k-i) \\ u(k+i+1|k-i) \\ \vdots \\ u(k+n_{\text{p}}|k-i) \\ \vdots \\ 0 \end{bmatrix}, \tag{5-3}$$

where $u(k+n_{\text{p}}|k-i)$ is the $(k+n_{\text{p}})$-th control input calculated at time $k-i$. At each new time step, the first control input is applied, and the remaining elements are shifted forward, with zeros appended at the end until all entries become zero.

Offers are only activated at the last possible moment, i.e. only if the offer is set to start at time $k+1$. Therefore, for the next time step, the shifted control sequence is updated according to

$$\boldsymbol{u}(k|k-1) = \begin{bmatrix} u(k+1|k-1) \\ \vdots \\ u(k+n_{\mathrm{p}}|k-1) \\ 0 \end{bmatrix}, \text{if } u(k|k) \neq 0, \boldsymbol{u}(k|k-1) = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \text{otherwise.} \tag{5-4}$$

This guarantees that the most recent data is used to make the activation decision.

The system dynamics (2-12b) are represented by the linearised transmission network model given in (4-5). The state constraint (2-12c) corresponds to the thermal limit on apparent power flows as defined in (4-9). However, since these are the only quadratic constraints, the overall computational burden can be significantly reduced by employing a linear approximation, effectively transforming the problem from a mixed-integer second-order cone program into a mixed-integer linear program. For systems operating at a high power factor (>0.95), it is reasonable to assume that the real power is approximately equal to the apparent power, i.e. , $P^{(n,m)}(k+i) \approx S^{(n,m)}(k+i)$. This assumption is well justified for high-voltage transmission networks, as many system operators mandate power factors above 0.95 to minimize line losses and improve operational efficiency [71]. Under this approximation, the constraint in (4-9) can therefore be reformulated as

$$|P^{(n,m)}(k+i)| < S^{\max}. \tag{5-5}$$

The state bounds (2-12d) enforce operational limits on nodal power injections. Finally, the input constraints (2-12e) are defined by the market activation and feasibility conditions specified in (4-17), (4-20), (4-21), (4-22), and (4-23).

For typical MPC formulations theoretical guarantees can be given on feasibility. However, since the control inputs are limited by market orders, guaranteeing satisfaction of the state constraints is not possible. To avoid infeasibility in the optimisation problem, a slack variable z is introduced to relax the constraints on power flows. This allows the optimisation to remain feasible even when strict constraint satisfaction cannot be achieved. The constraint limiting the power flow is then reformulated as

$$|P^{(n,m)}(k+i)| - S^{\max} \leq z^{(n,m)}(k+i) \qquad \forall i \in \{1,\dots,n_{\mathrm{p}}\}, (n,m) \in \mathscr{E}, \tag{5-6}$$

where the slack variable $z^{(n,m)}(k+i)$ represents the degree of constraint violation for line (n,m) at time $k+i$. To ensure that the optimisation problem does not arbitrarily increase the value of the slack variable, the following penalty term proportional to z is added to the objective function

$$\sum_{(n,m)\in\mathscr{E}} \sum_{i=1}^{n_{\mathrm{p}}} c_z z^{(n,m)}(k+i), \tag{5-7}$$

where the weighting coefficient c_z must be selected sufficiently large so that violating the constraint is always more costly than adjusting the control input, thereby maintaining the validity of the solution. Due to the high value of c_z it will be optimal to be as far as possible below the limit, i.e. $z^{(n,m)} < 0$, this incentives actions that reduce power even below the limit. To prevent this the following constraint is added on the slack variable

$$z^{(n,m)}(k+i) \geq 0 \quad \forall i \in \{1,\dots,n_{\mathrm{p}}\}. \tag{5-8}$$

This guarantees only constraint violation is penalised and as long as the use of the network stays within the constraints the power market stays unaffected. Combining all the equations and models results in the following definition for the MPC problem

$$
\begin{aligned}
\min_{\boldsymbol{x}(k+1),\boldsymbol{u}(k|k),\boldsymbol{\beta},\boldsymbol{\delta}} \; & \sum_{\mathrm{o}\in\mathcal{O}} m^{(\mathrm{o})} c^{(\mathrm{o})} \sum_{i=1}^{n_\mathrm{p}} \sum_{j=1}^{i} u^{(\mathrm{o})}(k+j|k) + \sum_{(n,m)\in\mathscr{E}} \sum_{i=1}^{n_\mathrm{p}} c_z z^{(n,m)}(k+i) \\
\text{s.t. } & \boldsymbol{x}(k+1) = S(k)\boldsymbol{x}(k) + T(k)\boldsymbol{u}(k|k) + T(k)\boldsymbol{w}(k) \\
& z^{(n,m)}(k+i) \geq 0 \quad \forall i \in \{1,\dots,n_\mathrm{p}\} \\
& |P^{(n)}(k+i)| < P^{(n),\max} \quad \forall i \in \{1,\dots,n_\mathrm{p}\}, n \in \mathscr{N} \\
& |Q^{(n)}(k+i)| < Q^{(n),\max} \quad \forall i \in \{1,\dots,n_\mathrm{p}\}, n \in \mathscr{N} \\
& |P^{(n,m)}(k+i)| - S^{\max} \leq z^{(n,m)}(k+i) \quad \forall i \in \{1,\dots,n_\mathrm{p}\}, (n,m) \in \mathscr{E} \\
& u^{(\mathrm{o})}(k+i) = \left[0 \dots 0 \; \Delta P_u^{(n),(\mathrm{o})}(k+i) \; 0 \dots 0\right]^T \quad \forall k+i \in \{t^{\mathrm{start},(\mathrm{o})},\dots,t^{\mathrm{stop},(\mathrm{o})}\}, \mathrm{o} \in \mathcal{O} \\
& \sum_{\mathrm{o}\in\mathcal{O}} u^{(\mathrm{o})}_{k+i|k} = 0 \quad \forall i \in \{0,\dots,n_\mathrm{p}-1\} \\
& \boldsymbol{u}(k) = \sum_{i=0}^{n_\mathrm{p}-1} \boldsymbol{u}(k|k-i) \\
& \beta^{\min,(\mathrm{o})} \leq \beta^{(\mathrm{o})} \leq 1 \\
\mathrm{PO}^{(\mathrm{o})}: & \Delta P_u^{(n),(\mathrm{o})}(k+i) \leq M^{\mathrm{high}} \delta^{(\mathrm{o})} - \sum_{j=t_\mathrm{start}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(j) \quad \forall i \in \{0,\dots,n_\mathrm{p}-1\} \\
& \Delta P_u^{(n),(\mathrm{o})}(k+i) \geq M^{\mathrm{low}} \delta^{(\mathrm{o})} - \sum_{j=t_\mathrm{start}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(k+j) \quad \forall i \in \{0,\dots,n_\mathrm{p}-1\} \\
& \Delta P_u^{(n),(\mathrm{o})}(k+i) \leq \beta^{(\mathrm{o})} P^{(n),(\mathrm{o})}(k+i) - \sum_{j=t_\mathrm{start}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(k+j) - M^{\mathrm{low}}(1-\delta^{(\mathrm{o})}) \\
& \qquad\qquad \forall i \in \{0,\dots,n_\mathrm{p}-1\} \\
& \Delta P_u^{(n),(\mathrm{o})}(k+i) \geq \beta^{(\mathrm{o})} P^{(n),(\mathrm{o})}(k+i) - \sum_{j=t_\mathrm{start}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(k+j) - M^{\mathrm{high}}(1-\delta^{(\mathrm{o})}) \\
& \qquad\qquad \forall i \in \{0,\dots,n_\mathrm{p}-1\} \\
\mathrm{FTO}^{(\mathrm{o})}: & \sum_{j=i}^{k+\ell_{\min}-1} \delta^{(\mathrm{o})}(j) \geq \ell_{\min}\left(\delta^{(\mathrm{o})}(k+i) - \delta^{\mathrm{o}}(k+i-1)\right) \quad \forall k+i \in \{t_\mathrm{start},\dots,t_\mathrm{stop}-\ell_{\min}+1\} \\
& \sum_{j=t_\mathrm{start}}^{t_\mathrm{stop}} \delta(j)^{(\mathrm{o})} \leq \ell^{\max} \\
& \sum_{j=i}^{t^{\mathrm{stop}}} \delta^{(\mathrm{o})}(j) \leq \ell^{\max}\left(1 + \delta^{(\mathrm{o})}(k+i) - \delta^{(\mathrm{o})}(k+i-1)\right) \forall k+i \in \{t^{\mathrm{start}} + \ell^{\min},\dots,t^{\mathrm{start}} + \ell^{\max} - 1\} \\
& \Delta P_u^{(n),(\mathrm{o})}(k+i) \leq M^{\mathrm{high}} \delta^{(\mathrm{o})}(k+i) - \sum_{j=t^{\mathrm{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(j) \quad \forall i \in \{0,\dots,n_\mathrm{p}-1\} \\
& \Delta P_u^{(n),(\mathrm{o})}(k+i) \geq M^{\mathrm{low}} \delta^{(\mathrm{o})}(k+i) - \sum_{j=t^{\mathrm{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(j) \quad \forall i \in \{0,\dots,n_\mathrm{p}-1\} \\
& \Delta P_u^{(n),(\mathrm{o})}(k+i) \leq \beta^{(\mathrm{o})} P^{\max,(\mathrm{o})} - \sum_{j=t^{\mathrm{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(j) - M^{\mathrm{low}}(1-\delta^{(\mathrm{o})}(k+i)) \\
& \qquad\qquad \forall i \in \{0,\dots,n_\mathrm{p}-1\} \\
& \Delta P_u^{(n),(\mathrm{o})}(k+i) \geq \beta^{(\mathrm{o})} P^{\max,(\mathrm{o})} - \sum_{j=t^{\mathrm{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(j) - M^{\mathrm{high}}(1-\delta^{(\mathrm{o})}(k+i)) \\
& \qquad\qquad \forall i \in \{0,\dots,n_\mathrm{p}-1\}
\end{aligned}
\tag{5-9}
$$

,

where the constraints for $\text{FTO}^{(\text{o})}$ and $\text{PO}^{(\text{o})}$ are repeated for every order in their respective sets using the parameters defined in tuple of the offer. where the constraints associated with $\text{FTO}^{(\text{o})}$ and $\text{PO}^{(\text{o})}$ are applied for every offer o in their respective sets, using the parameters specified in each offer's tuple. The binary activation variables are collected in the vector $\boldsymbol{\delta}$, and the corresponding scaling coefficients in vector $\boldsymbol{\beta}$. The aggregated control input at time k is given by $\boldsymbol{u}(k) = \sum_i \boldsymbol{u}(k|k-i)$, which represents the cumulative effect of all control actions scheduled at previous time steps that remain active at time k.

5-3 CM as an CC-MPC problem

The MPC formulation presented in the previous subsection can be extended to a sample approximated CC-MPC by replacing the dynamics with

$$\boldsymbol{x}^{(\text{s})}(k+1) = S(k)x(k) + T(k)\boldsymbol{u}(k) + T(k)\boldsymbol{w}^{(\text{s})}(k) \quad \forall \text{s} \in \{1, \dots, n_{\text{s}}\} \tag{5-10}$$

where $\boldsymbol{x}^{(\text{s})}(k+1)$ is a the state evolution over the full prediction horizon under scenario s and n_{s} is the number of scenarios. The deterministic constraint in the MPC formulation must be replaced with the scenario approximation of the stochastic constraint, which can be formulated as

$$\frac{1}{n_{\text{s}}} \sum_{s=1}^{n_{\text{s}}} \mathbb{1}\left(|P^{(n,m),(\text{s})}(k+i)| - S^{\max} \geq 0\right) \leq 1 - \alpha \tag{5-11}$$

where $\mathbb{1}(\cdot)$ is defined as in (2-16) meaning if the constraint is met the indicator function becomes 1 and if the constraint is violated it is 0. The risk parameter $\alpha \in [0, 1)$ defines the minimum fraction of scenarios that satisfy the limits. Similarly to the MPC formulation the constraints have to be replaced by a penalty since constraint satisfaction cannot be guaranteed. The indicator function is replaced by binary variables $\delta^{(\text{s})}$ and the slack variables $z^{(n,m)}(k+i)$ are re-introduced in the following way

$$|P^{(n,m),(\text{s})}(k+i)| - S^{\max} \leq z^{(n,m)}(k+i) + M\delta^{(\text{s})} \quad \forall i \in \{1, \dots n_{\text{p}}\}, \text{s} \in \{1, \dots, n_{\text{s}}\}, \tag{5-12}$$

where $z^{(n,m)}(k+i)$ is the shared over all scenarios and thus reflects the maximum constraint violation for transmission line (n, m) at time $k+i$ over all the scenarios. The penalty in the objective function is then formulated as

$$c_{\text{s}} \max\left(\sum_{s=1}^{n_{\text{s}}} \delta^{(\text{s})} - (1-\alpha)n_{\text{s}}, 0\right) + \sum_{(n,m)\in\mathscr{E}} \sum_{i=1}^{n_{\text{p}}} c_z z^{(n,m)}(k+i) \tag{5-13}$$

where c_{s} must be selected such that additional scenarios exhibiting constraint violations incur a penalty greater than the magnitude of the violation itself. This allows the slack variable $z^{(n,m)}(k+i)$ to stay zero in the scenarios where constraint violation is 'allowed' due to the chance constraints. This combination penalises the magnitude of constraint violations in the scenarios that must be satisfied, but excludes the constraint violations in the scenarios where constraint violation is allowed in due to the risk parameter in the chance-constraints. Combining all gives us the following novel formulation for a market-based CC-MPC CM

method

$$
\begin{aligned}
\min_{x_k,\boldsymbol{u}_{k|k},\boldsymbol{\beta},\boldsymbol{\delta}} \; & \sum_{\mathrm{o}\in\mathcal{O}} m^{(\mathrm{o})}c^{(\mathrm{o})}\sum_{i=1}^{n_\mathrm{p}}\sum_{j=1}^{i}u^{(\mathrm{o})}(k+j) + c_\mathrm{s}\max\left(\sum_{s=1}^{n_\mathrm{s}}\delta^{(\mathrm{s})} - (1-\alpha)n_\mathrm{s}, 0\right) + \sum_{(n,m)\in\mathscr{E}}\sum_{i=1}^{n_\mathrm{p}} c_z z^{(n,m)}(k+i) \\
\text{s.t. } & \boldsymbol{x}^{(\mathrm{s})}(k) = S(k)x(k) + T(k)\boldsymbol{u}(k) + T(k)\boldsymbol{w}^{(\mathrm{s})}(k|k) \quad \forall \mathrm{s}\in\{1,\dots,n_\mathrm{s}\} \\
& |P^{(n)}(k+i)| < P^{(n),\max} \quad \forall i\in\{0,\dots,n_\mathrm{p}\}, n\in\mathscr{N} \\
& |Q^{(n)}(k+i)| < Q^{(n),\max} \quad \forall i\in\{0,\dots,n_\mathrm{p}\}, n\in\mathscr{N} \\
& |P^{(n,m),(\mathrm{s})}(k+i)| - S^{\max} \le z^{(n,m)}(k+i) + M^{\mathrm{high}}\delta^{(\mathrm{s})} \quad \forall i\in\{1,\dots n_\mathrm{p}\}, \mathrm{s}\in\{1,\dots,n_\mathrm{s}\} \\
& u^{(\mathrm{o})}(k+i) = \begin{bmatrix}0\dots 0\ \Delta P_u^{(n)}(k+i)\ 0\dots 0\end{bmatrix}^T \quad \forall k+i\in\{t^{\mathrm{start},(\mathrm{o})},\dots,t^{\mathrm{stop},(\mathrm{o})}\}, \mathrm{o}\in\mathcal{O} \\
& \sum_{\mathrm{o}\in\mathcal{O}} u^{(\mathrm{o})}_{k+i|k} = 0 \quad \forall i\in\{0,\dots,n_\mathrm{p}-1\} \\
& \boldsymbol{u}(k) = \sum_{i=0}^{n_\mathrm{p}} \boldsymbol{u}(k|k-i) \\
& \beta^{\min,(\mathrm{o})} \le \beta^{(\mathrm{o})} \le 1 \\
\mathrm{PO}^{(\mathrm{o})}: & \Delta P_u^{(n),(\mathrm{o})}(k+i) \le M^{\mathrm{high}}\delta^{(\mathrm{o})} - \sum_{j=t_\mathrm{start}}^{k+i-1}\Delta P_u^{(n),(\mathrm{o})}(j) \forall i\in\{0,\dots,n_\mathrm{p}-1\} \\
& \Delta P_u^{(n),(\mathrm{o})}(k+i) \ge M^{\mathrm{low}}\delta^{(\mathrm{o})} - \sum_{j=t_\mathrm{start}}^{k+i-1}\Delta P_u^{(n),(\mathrm{o})}(k+j) \forall i\in\{0,\dots,n_\mathrm{p}-1\} \\
& \Delta P_u^{(n),(\mathrm{o})}(k+i) \le \beta^{(\mathrm{o})}P^{(n),(\mathrm{o})}(k+i) - \sum_{j=t_\mathrm{start}}^{k+i-1}\Delta P_u^{(n),(\mathrm{o})}(k+j) - M^{\mathrm{low}}(1-\delta^{(\mathrm{o})}) \forall i\in\{0,\dots,n_\mathrm{p}-1\} \\
& \Delta P_u^{(n),(\mathrm{o})}(k+i) \ge \beta^{(\mathrm{o})}P^{(n),(\mathrm{o})}(k+i) - \sum_{j=t_\mathrm{start}}^{k+i-1}\Delta P_u^{(n),(\mathrm{o})}(k+j) - M^{\mathrm{high}}(1-\delta^{(\mathrm{o})}) \forall i\in\{0,\dots,n_\mathrm{p}-1\} \\
\mathrm{FTO}^{(\mathrm{o})}: & \sum_{j=i}^{k+\ell_{\min}-1}\delta^{(\mathrm{o})}(j) \ge \ell_{\min}(\delta^{(\mathrm{o})}(k+i) - \delta^{\mathrm{o}}(k+i-1)) \quad \forall k+i\in\{t_\mathrm{start},\dots,t_\mathrm{stop}-\ell_{\min}+1\} \\
& \sum_{j=t_\mathrm{start}}^{t_\mathrm{stop}}\delta(j)^{(\mathrm{o})} \le \ell^{\max} \\
& \sum_{j=i}^{t^{\mathrm{stop}}}\delta^{(\mathrm{o})}(j) \le \ell^{\max}(1+\delta^{(\mathrm{o})}(k+i) - \delta^{(\mathrm{o})}(k+i-1)) \\
& \qquad\qquad \forall k+i\in\{t^{\mathrm{start}}+\ell^{\min},\dots,t^{\mathrm{start}}+\ell^{\max}-1\} \\
& \Delta P_u^{(n),(\mathrm{o})}(k+i) \le M^{\mathrm{high}}\delta^{(\mathrm{o})}(k+i) - \sum_{j=t^{\mathrm{start}}}^{k+i-1}\Delta P_u^{(n),(\mathrm{o})}(j) \quad \forall i\in\{0,\dots,n_\mathrm{p}-1\} \\
& \Delta P_u^{(n),(\mathrm{o})}(k+i) \ge M^{\mathrm{low}}\delta^{(\mathrm{o})}(k+i) - \sum_{j=t^{\mathrm{start}}}^{k+i-1}\Delta P_u^{(n),(\mathrm{o})}(j) \quad \forall i\in\{0,\dots,n_\mathrm{p}-1\} \\
& \Delta P_u^{(n),(\mathrm{o})}(k+i) \le \beta^{(\mathrm{o})}P^{\max,(\mathrm{o})} - \sum_{j=t^{\mathrm{start}}}^{k+i-1}\Delta P_u^{(n),(\mathrm{o})}(j) - M^{\mathrm{low}}(1-\delta^{(\mathrm{o})}(k+i)) \quad \forall i\in\{0,\dots,n_\mathrm{p}-1\} \\
& \Delta P_u^{(n),(\mathrm{o})}(k+i) \ge \beta^{(\mathrm{o})}P^{\max,(\mathrm{o})} - \sum_{j=t^{\mathrm{start}}}^{k+i-1}\Delta P_u^{(n),(\mathrm{o})}(j) - M^{\mathrm{high}}(1-\delta^{(\mathrm{o})}(k+i)) \quad \forall i\in\{0,\dots,n_\mathrm{p}-1\}
\end{aligned}
\tag{5-14}
$$

,

5-4 CM as an CC-MPC-RL problem

The proposed method employs RL to enhance the performance of the CC-MPC formulation in (5-14). The RL-agent learns to dynamically adjust a scaling factor for the variance of the

innovations in the SARIMA-based disturbance prediction models. The concept is inspired by the adaptive robustification factor κ introduced by [72], where a deep reinforcement learning agent continuously adjusts κ to balance feasibility and performance in an adaptive stochastic non-linear MPC framework. Similar to their approach, the RL-agent here learns to modify the uncertainty scaling factor online based on most recent measurements, allowing the controller to respond more effectively to time-varying uncertainties present in the system.

An overview of the proposed control architecture is presented in Figure 5-1. The 'true' system generates the state feedback $x(k)$ and disturbance signal $w(k)$, which are passed to the ARMA models for disturbance prediction. The RL-agent observes the system's behaviour and past control performance, then determines an appropriate scaling factor κ_b for each of the nodes with uncertainty, which adjusts the variance of the predicted disturbance distribution $w(k+i)$. This prediction is then provided to the CC-MPC, which computes the optimal control action $u(k)$. Once applied to the system, the process repeats at each sampling instant, forming a closed control loop. In this configuration, the RL-agent operates in parallel with the predictive control framework, continuously improving its policy for adaptive uncertainty handling. In

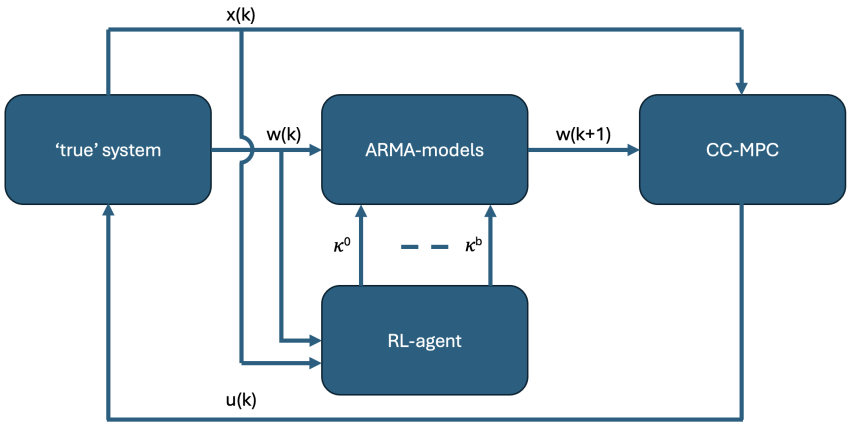


Figure 5-1: Overview of the proposed closed-loop framework integrating RL with CC-MPC. The RL-agent dynamically adjusts the innovation variance in the ARMA-based disturbance predictions to improve robustness and adaptability.

the following subsections, the full derivation of the RL-agent will be presented starting with the necessary definitions of state space, actions space, etc. and will finish with the training procedure.

State space

The state space of the RL agent consists of the system state vector $x(k)$ and the previous 192 values (corresponding to two days of data) of the disturbance state vector $w(k)$ as defined in (4-2). The RL state vector is therefore defined as

$$s(k) = \begin{bmatrix} x(k)^T & w(k)^T & w(k-1)^T & \dots & w(k-192)^T \end{bmatrix}^T \in \mathbb{R}^{n_x+192n_w}, \tag{5-15}$$

where n_x and n_d denote the dimensions of the system and disturbance states, respectively. The system dynamics evolve according to the plant model

$$x(k+1) = A(k)x(k) + B(k)u(k) + B(k)w(k), \tag{5-16}$$

while the disturbance $w(k)$ follows the SARIMA dynamics,

$$w(k+1) = \sum_{i=1}^{p} \phi_i \, w(k+1-i) + \sum_{j=1}^{q_s} \Theta_j \, \varepsilon(k+1-j) + \varepsilon(k+1), \quad \varepsilon(k+1) \sim \mathcal{N}(0, \sigma^2). \tag{5-17}$$

Action space

The action space represents the possible decisions available to the RL agent at each time step. Here, an action corresponds to a scaling factor applied to the variance of the innovations in the SARIMA disturbance models. For each of the eight regions, the variance of the innovation term is scaled by κ_b, such that

$$\varepsilon(k+1) \sim \mathcal{N}(0, \, \kappa_b \sigma^2), \tag{5-18}$$

where κ_b adjusts the uncertainty level of the forecast. The full action set is therefore

$$\begin{aligned} \mathcal{A} &= [\kappa_1, \ldots, \kappa_8], \\ \text{where} \quad \kappa_b &\in \{0.5, \, 0.6, \, \ldots, \, 2.0\}, \quad \forall b \in \{1, \ldots, 8\}. \end{aligned}$$

Each action corresponds to a joint assignment of eight scaling factors, one per SARIMA model. Since each κ_b can take 16 discrete values, the total number of possible joint actions is 16^8. This exponential growth makes standard Deep Q-Network (DQN) approaches computationally infeasible, as exhaustive exploration of all combinations is impractical. To overcome this challenge, an action-branching architecture is adopted, as described in Subsection 5-4.

Reward function

The reward function transforms the control objective into a scalar feedback signal. In this formulation, it balances the trade-off between conservative and risk-seeking actions by combining operational costs with constraint violations. The immediate reward at time step t is given by

$$\begin{aligned} r(k) &= R(s(k), a(k), s(k+1)) \\ R(s(k), a(k), s(k+1)) &= -c_{\text{cost}} C_{\text{cost}}(a(k), s(k)) - c_{\text{vio}} C_{\text{vio}}(s(k+1)) \end{aligned} \tag{5-20}$$

where c_{cost}, c_{vio} are a scaling factors of the cost and limit violation for training stability, $C(a(k), s(k))$ is the financial cost of the actions taken by the CC-MPC as a result of $a(k)$ as defined in (4-24), and $C_{\text{vio}}(s(k+1))$ is the of constraint violation with as the square of the violation, defined as

$$C_{\text{vio}}(s(k+1)) = \sum_{(n,m) \in \mathscr{E}} \max(0, |P^{(n,m)}(k+1)| - S^{\max})^2. \tag{5-21}$$

Policy

The policy $\pi(a|s)$ defines the agent's strategy, mapping states $s \in \mathcal{S}$ to actions $a \in \mathcal{A}$. For a set policy π the value of an action $a(k)$ for state $s(k)$ is determined using the state-value function $Q_\pi(s(k), a(k))$, which is defined as follows [73]:

$$Q_\pi(s(k), a(k)) = \mathbb{E}\Big[\sum_{k=0}^{\infty} \gamma^k R(s(k), a(k), s(k+1)) | s = s(k), a(k) = a, \pi\Big] \tag{5-22}$$

where Q_π is the action-value function; γ the discount factor; $r(k)$ is the reward achieved by taking action $a(k)$,; and $a(k)$, $s(k)$ and π are the current action, state and policy respectively. The action-value function $Q(s(k), a(k))$ thus says what the expected total value of the action will be given that after that action $a(k)$ it always takes the action that follows the policy π. The optimal action-value $Q^*(s(k), a(k))$ is then achieved by finding the policy that maximises the expected value of the reward

$$Q^*_\pi(s(k), a(k)) = \max_\pi \mathbb{E}\Big[\sum_{k=0}^{\infty} \gamma^k R(s(k), a(k), s(k+1)) \Big| s = s(k), a = a(k), \pi\Big]. \tag{5-23}$$

The optimal action-value function follows the structure of the Bellman equation, which originates from dynamic programming. The Bellman equation decomposes the decision problem into two parts: the expected immediate reward of the current action and the expected reward of the future decisions. This recursive structure enables the formulation of optimal policies as an iterative problem. The optimal action-value function can be expressed as [73]:

$$Q^*(s(k), a(k)) = \mathbb{E}\Big[r(k) + \gamma \max_{a(k+1)} Q^*(s(k+1), a(k+1)) \Big| s = s(k), a = a(k)\Big]. \tag{5-24}$$

Consequently, the problem shifts from directly determining the optimal policy to learning the action-value function, which simultaneously captures the immediate reward of an action and the expected return of the remaining decision process. Typically for discrete state spaces it is not feasible to learn exact values for all state-action pairs, therefore the goal of RL is to approximate this action-value function as follows [73, 74]

$$Q(s(k), a(k); \theta) \approx Q^*(s(k), a(k)), \tag{5-25}$$

where θ are the parameters of the function approximator. In the specific case of deep Q-learning or DQN the function approximator is an neural network where the weights θ are the parameters of the neural network. The parameters are updated using the difference between temporal difference target $y(k)$ and the current estimate of the action-value $Q(s(k), a(k); \theta)$, using this recursive formulation from the Bellman equation. The temporal difference target is defined as

$$y(k) = r(k) + \gamma \max_{a(k+1)} Q(s(k+1), a(k+1)), \tag{5-26}$$

and t temporal difference error is then be defined as follows:

$$\mathrm{TD}(k) = y(k) - Q(s(k), a(k); \theta(k)). \tag{5-27}$$

In DQN the squared temporal difference error is used as loss function to update the parameters $\theta(k)$ of the network, as follows:

$$L(\theta(k)) = \mathbb{E}\big[(y(k) - Q(s(k), a(k); \theta(k)))^2\big]. \tag{5-28}$$

It is easy to see that if $\mathrm{TD}(k) = 0$ the learned action-value function perfectly predicts the action-value, and thus can be used to choose the optimal action given the current state.

DQN architecture

As discussed in Subsection 5-4, the number of possible joint actions grows exponentially with the number of uncertain nodes. To address this, an action-branching DQN architecture is employed. This approach decomposes the high-dimensional joint action space into multiple independent branches—one for each node with uncertainty—while maintaining a shared state representation. Instead of evaluating all possible combinations of actions jointly, each branch independently estimates the advantage of its own local action, and these are later combined with the shared state value to determine the overall Q-value. This decomposition allows the computational complexity to scale linearly with the number of nodes rather than exponentially, since the network learns a separate policy for each branch instead of a single high-dimensional joint policy [75]. The architecture integrates action branching with double Q-learning, experience replay, and a duelling network structure. These components are discussed first, followed by a detailed explanation of the full method used in this work.

Experience replay

The use of experience replay has improved training stability of RL [73]. Instead of updating the Q-network solely from the most recent transition, the agent stores past interactions $e(k) = (s(k), a(k), r(k), s(k+1))$ in a replay buffer $\mathcal{D} = \{e_1, e_2, \ldots, e(k)\}$. During training, mini-batches of experiences are drawn uniformly at random from this buffer, which breaks correlations between consecutive samples and improves data efficiency. The interaction in Figure 2-4 is then adapted to Figure 5-2 where instead of directly returning the state, action, reward and next state they are saved in a the replay memory and randomly sampled to update the network parameters each time step.

Double Q-learning

Standard Q-learning uses the same network for action selection and evaluation, this is one of the reasons that standard Q-learning tends to overestimate action values. Double Q-learning addresses this issue by using two networks: one to select the best action, the policy (or online) network $Q(\cdot)$, and another to evaluate the chosen action, the target network $Q'(\cdot)$ [74]. This decoupling reduces overestimation bias by separating action selection from evaluation. Then every τ update steps the parameters from the policy network are copied into the target network. Formally, the Double Q-learning temporal difference target is defined as:

$$y(k) = r(k+1) + \gamma Q'\Big(s(k+1), \arg\max_{a(k+1)} Q(s(k), a(k+1); \theta); \theta'\Big), \tag{5-29}$$

where θ are the parameters of the policy network (used for action selection), and θ' are the parameters of the target network (used for action evaluation). The architecture of each model does not change but the overall architecture does as depicted in Figure 5-3. Instead of one model two neural networks are used to evaluate the actions as defined in (5-29).

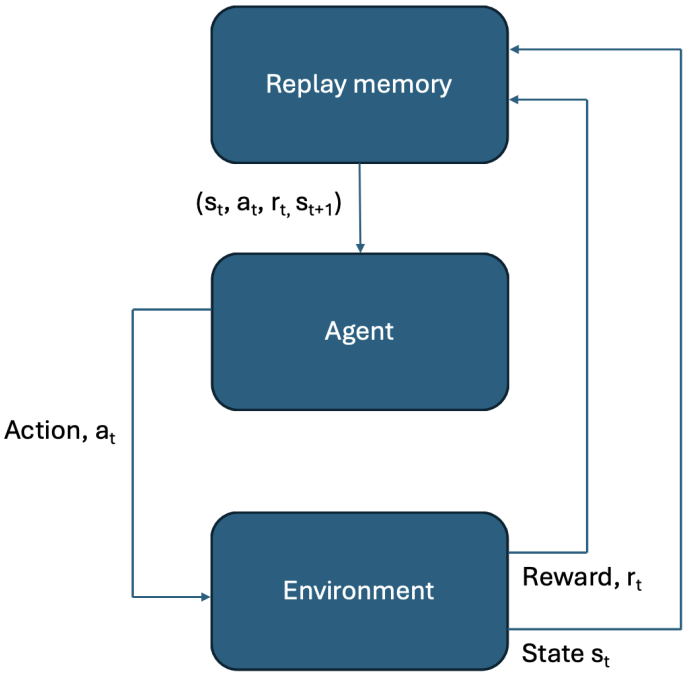


Figure 5-2: Experience replay

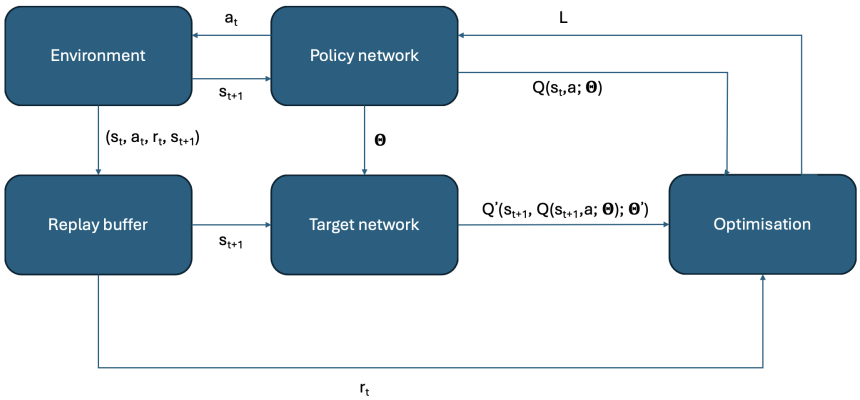


Figure 5-3: Double Q-learning architecture

Duelling

The duelling network architecture modifies the standard DQN by decomposing the Q-function into two separate estimators: the state-value function $V(\mathrm{s})$ and the state-dependent advantage function $A(s, a)$. Intuitively, in many states the choice of action has little effect on outcomes, so estimating the value of the state directly is more efficient than learning separate Q-values for every action. The state-value function $V(s(k))$ is very similar to the action-value function $Q(s(k), a(k))$ but instead of taking a action that is not defined by the policy first and then

following the policy it follows the policy immediately, this can be formalised as follows [76]:

$$\begin{aligned} V_\pi(s(k)) &= \mathbb{E}_{a(k)=\pi(s)}\left[Q_\pi(s(k),a(k))\right] \\ &= \mathbb{E}\left[\sum_{k=0}^{\infty}\gamma^k r(k)|s=s(k),\pi\right]. \end{aligned} \tag{5-30}$$

The values for the state and each action are then combined into the action-advantage function $A(s(k),a(k))$:

$$\begin{aligned} A(s(k),a(k)) &= V(s(k)) - Q(s(k),a(k)) \\ Q(s(k),a(k)) &= V(s(k)) + \left(A(s(k),a(k)) - \max_{a'(k)} A(s(k),a(k)')\right). \end{aligned} \tag{5-31}$$

This adaptation also changes the architecture of the network. Instead of only sequential layers that stack until the output layer of the Q-values, the duelling network introduces has layers that are parallel but share the same shared feature extraction layers: one estimating the scalar state-value function $V(s(k))$, and another estimating the advantage function $A(s(k),a(k))$. These two outputs are then combined in an aggregation layer to produce the final Q-values for all actions, as illustrated in Figure 5-4.

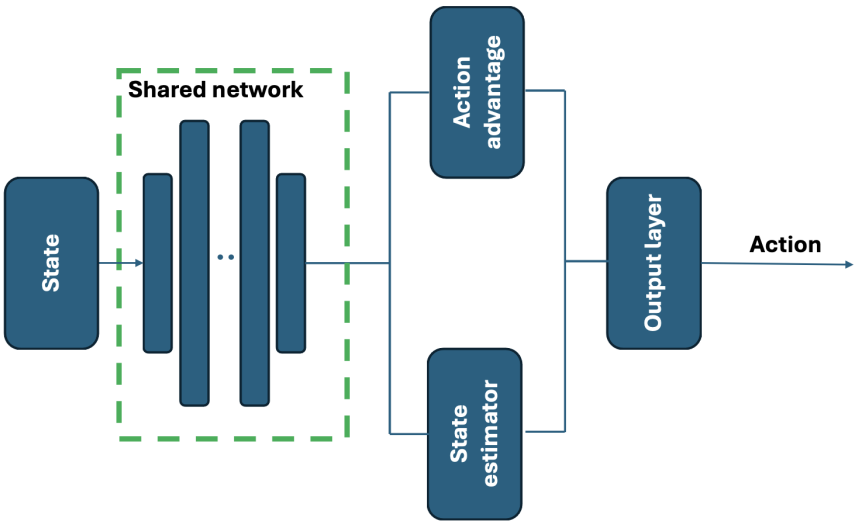


Figure 5-4: Duelling network architecture

Formally, the duelling architecture defines the Q-function as

$$Q(s(k),a(k);\theta,\alpha,\beta) = V(s(k);\theta,\beta) + \left(A(s(k),a(k);\theta,\alpha) - \max_{a'(k)} A(s(k),a';\theta,\alpha)\right), \tag{5-32}$$

where θ are the parameters of the shared feature extraction layers and aggregation layer, and α and β are the parameters of the parallel advantage and value layers, respectively. This decomposition improves learning stability and efficiency by allowing the network to better estimate state values, particularly in situations where the choice of action has little influence on the outcome [76].

Action branching framework

The overall network architecture employed in this work is based on the action branching architecture proposed in [75]. An illustration of the network is shown in Figure 5-5. The key idea is to split the network into multiple branches, in this case 8, instead of one. In this way each branch can select one action instead of selecting the best combination of 8 actions.

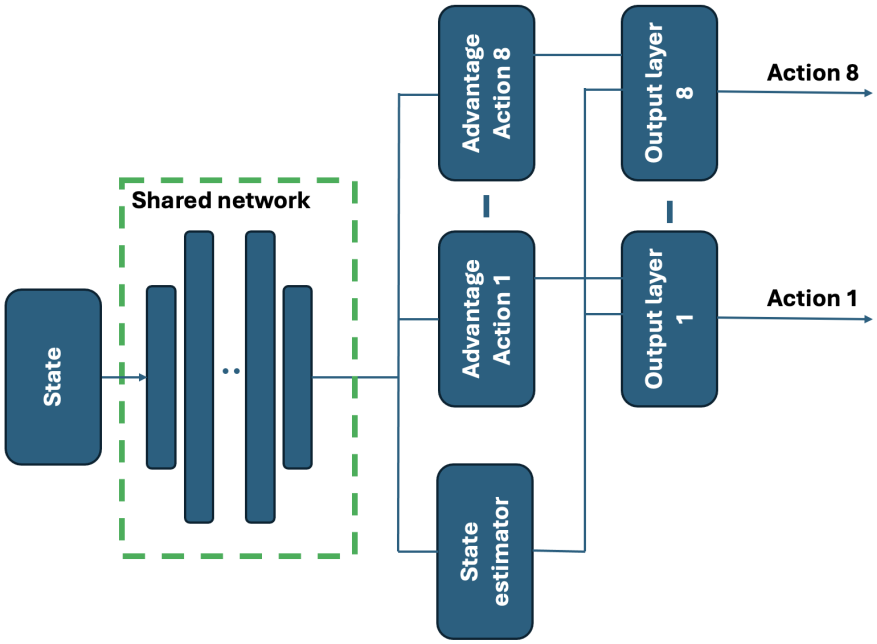


Figure 5-5: Action branching architecture

This approach is combined with the innovation of the double deep Q-learning (Figure 5-3 and experience replay (Figure 5-2). Each branch of the network outputs action-specific advantage values for one action dimension, while a shared representation (the state-value branch) estimates the state-value function, following the principles of duelling DQN [76]. This shared state-value captures the common utility of being in a given state, while the branch-specific advantages model the relative value of selecting a particular action within that branch. The combination of these streams results in an action-value function that respects the duelling decomposition [75]:

$$Q^{(b)}(s(k),a(k)^{(b)}) = V(s(k)) + A^{(b)}(s(k),a^{(b)}(k)) - \frac{1}{|\mathcal{A}^{(b)}|}\sum_{a(k)^{(b)}\in\mathcal{A}^{(b)}} A^{(b)}(s(k),a^{(b)}(k)) \quad \forall b\in\{1,\dots,8\}, \tag{5-33}$$

where b the denotes the branch, $V(s(k))$ is the shared state-value, and $A^{(b)}(s(k),a(k)^{(d)})$ denotes the action-advantage for an action in branch b. In a similar fashion, but according to the double Q-learning update, the temporal difference target needs to be changed for each branch

$$y^{(b)}(k) = r(k+1) + \gamma Q'^{(b)}\left(s(k+1), \arg\max_{a^{(b)}(k+1)} Q^{(b)}(s(k+1),a^{(b)}(k+1);\theta);\theta'\right), \tag{5-34}$$

where $Q^{(b)}$ and $Q'^{(b)}$ denote for branch b of the policy and target network respectively. The loss function is then defined as the square of the mean squared temporal difference error as follows

$$L(\theta(k)) = \mathbb{E}\left[\frac{1}{8}\sum_{b=1}^{8}\left(y^{(b)}(k) - Q^{(b)}(s(k), a^{(b)}(k); \theta(k))\right)^2\right], \tag{5-35}$$

as this has been shown to produce the best results [75].

Training procedure

The full training procedure for the RL agent is summarized in Algorithm 1. At each time step, the current state $s(k)$, which includes both the measured system state and recent disturbances as defined in (5-15), is observed by the agent. Based on its exploration strategy it selects an action $a(k)$ the best action according to the current policy with probability $1-\epsilon$ or samples an action at random with probability ϵ. With one action consisting of 8 uncertainty scaling factors $\kappa_b(k)$ discussed in Subsection 5-4. These scaling factors adjust the variance of the disturbance predictions generated by the SARIMA model as defined in (5-18), the resulting predictions are then used in the CC-MPC problem in (5-14) to compute the optimal control sequence.

The resulting control input $u(k)$ is applied to the 'true system', producing the next state $s(k+1)$. The reward $r(k)$ is computed using (5-20), which penalizes both the operational cost of $u(k)$ and any constraint violations. The experience tuple $(s(k), a(k), r(k), s(k+1))$ is then stored in the replay buffer to break correlations between sequential time steps and accelerate learning. During training, batches of these experiences are randomly sampled from the buffer to update the neural network parameters.

The loss is computed using the branch-based temporal difference error defined in (5-35), while double Q-learning is used to reduce overestimation bias and the duelling network architecture improves value estimation stability. Periodic updates, every τ time steps, of the target network further enhance training robustness. Through this iterative closed-loop process, the RL-agent continuously refines its policy for online uncertainty adaptation, enabling the CC-MPC to achieve a better performance under dynamically varying disturbances.
The results of the training procedure are shown in Figure 5-6. Where each episode is a full day of simulation and the light red line represents the raw cumulative reward obtained in each training episode, while the darker red line shows a smoothed average to highlight overall trends .Throughout the training process, the cumulative reward exhibits oscillatory behaviour, with frequent fluctuations and several deep negative spikes. This behaviour indicates that the learning process is unstable, which can be explained by the fact that RL is penalised for constraint violation while in most cases it is impossible to completely remove the congestion. The absence of a clear upward trend suggests that the agent does not converge towards a consistently improved policy. Instead, the cumulative rewards oscillate around a relatively constant mean value, showing the relative inability of the controller (since the reward is largely determined by which day is sampled to train on) to influence the reward.

Algorithm 1: Training of the RL–enhanced CC–MPC controller

Init: Initialize Q_θ, $Q_{\bar{\theta}} \leftarrow Q_\theta$, replay buffer $\mathcal{D}$

while *training* **do**

$s(k) \leftarrow [x(k)^T \;\; w_k^T \ldots w_{k-192}^T]^T$

if $Unif(0,1) > \varepsilon(k)$ **then**

$a(k) \leftarrow \arg\max_a Q_\theta(s(k), a)$

else

$\kappa_b \sim \text{UNIF}(\{0.5, 0.6, \ldots, 2.0\}) \quad \forall b \in \{1, \ldots, 8\}$

end

$\kappa_b \leftarrow \arg\max_{a^{(b)}(k)} Q^{(b)}(s(k), a^{(b)}(k)) \quad \forall b \in \{1, \ldots, 8\}$

$\boldsymbol{w}(k+1) \leftarrow \text{SARIMA}^{(b)}(\kappa(k)^{(b)})$

$\boldsymbol{u}(k) \leftarrow$ Solution of (5-14)

$x(k+1) \leftarrow \text{True system}(x(k), u(k), w(k))$

$s(k+1) = [x^T(k+1)\, w^T(k+1)\, w^T(k)\, w^T(k-1) \;\ldots\; w^T(k-191)]^T$

$r(k) \leftarrow R(s(k), a(k), s(k+1))$

Store $(s(k), a(k), r(k), s(k+1))$ in $\mathcal{D}$

$\{(s(k), a(k), r(k), s(k+1))\}_{=1}^{\text{batchsize}} \leftarrow \mathcal{D}$

$L(\theta(k)) = \mathbb{E}\left[\frac{1}{n_d} \sum_b \left(y^{(b)}(k) - Q^{(d)}(s(k), a^{(b)}(k); \theta(k))\right)^2\right]$,

Update $\theta(k)$ with gradient step on $L(\theta(k))$ $k \leftarrow k+1$ Every τ steps: $\theta(k)' \leftarrow \theta(k)$

end

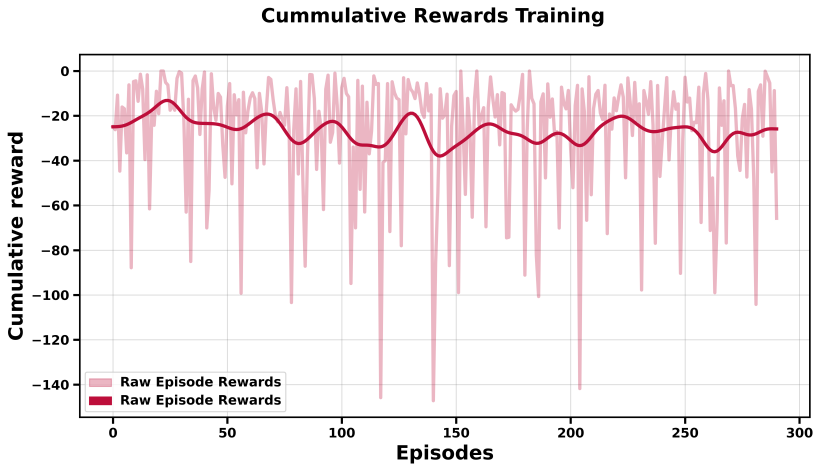


Figure 5-6: Cumulative rewards during reinforcement learning training.

5-5 Summary

This chapter presents control methods for CM, moving from a fast heuristic to optimization- and learning-based controllers. It starts with a physics-validated greedy matching baseline that resolves predicted congestion by activating the cheapest feasible pair of market offers. Next, CM is posed as an economic, market-based MPC that respects the persistence of

activated offers over the horizon. The method is then extended to a sample-approximated CC-MPC that enforces chance constraints under uncertainty. Finally, the proposed controller integrates RL with CC-MPC, where an action-branching, duelling, double DQN agent adapts scenario uncertainty scaling per node to balance cost and reliability

Chapter 6

Case study and results

In this chapter, the results of the methods developed in Chapter 5 are presented and compared. First, the simulation and experimental setups are described, providing the necessary context for how the results were obtained. Next, a simulation on the Dutch high voltage grid is discussed as a detailed case study. Subsequently, three key aspects are analysed: the impact of prediction quality on controller performance, the influence of the risk parameter parameter on the proposed CC-MPC method, and the effect of the RL-based enhancement of the MPC. Finally, the overall results are summarised and compared with other existing CM approaches.

6-1 Simulation and experimental setup

This study evaluates the performance of the proposed control strategy under varying modelling and operational conditions. To assess the effect of the prediction horizon on model performance, two different lengths are considered. The prediction horizon defines how far ahead the controller anticipates system behaviour, and its selection involves a fundamental trade-off. A longer horizon enables the inclusion of longer admissible offers, providing greater flexibility for market participants and enhancing the overall feasibility of offer matching. However, it may also increase the accumulation of model mismatch, particularly for linearised system representations that are valid only near the current operating point, and it leads to higher computational complexity during optimisation. To study the influence of the risk parameter α, two values of α are used. This parameter represents the level of risk aversion in the decision-making process, affecting how conservative or aggressive the control actions are. To examine the impact of prediction accuracy, the MPC is tested using forecasts generated by the SARIMA models developed in Chapter 3. This is then compared with the MPC scheme that uses perfect predictions (MPC-PP) of the future uncertainty values. This comparison shows the effect of forecast errors on control performance. Finally, to assess the effect of offer quantity; four different offer sets are analysed. Each set represents a distinct configuration of available options two for each prediction horizon length with two different amounts of offers, of allowing evaluation of how market flexibility influences outcomes.

In the following sections, the control schemes outlined in the previous chapters, namely the Algorithmic Greedy Matching approach, the MPC, the CC-MPC, and the RL enhanced CC-MPC, are evaluated. The objective is to validate these strategies in terms of constraint satisfaction and congestion cost. To this end, the influence of several key design choices is investigated: the prediction horizon, the safety level α, and the market flexibility. Specifically, two prediction horizon lengths are considered to analyse how the look-ahead period affects performance. A longer horizon enables the inclusion of longer admissible offers, providing greater flexibility for market participants and increasing the overall feasibility of offer matching. However, it may also increase the accumulation of model mismatch, particularly for linearised system representations that are valid only near the current operating point, and it leads to higher computational complexity during optimisation. The safety level α controls the degree of risk aversion in the chance constraints, determining how conservative or aggressive the control actions are; two values of α are tested to evaluate its impact on system performance.

To examine the effect of forecast quality, the MPC is tested under two conditions: one using the stochastic forecasts generated by the SARIMA models introduced in Section 3, and another using perfect predictions (MPC-PP), which uses to the actual future values. This comparison quantifies the influence of forecast errors on control performance. Finally, the sensitivity to market flexibility is analysed by varying the number of available offers, 100 and 200 per horizon, across both prediction horizon lengths. For benchmarking purposes, all methods are compared against two reference cases: (i) an uncontrolled scenario, where no congestion management actions are applied, and (ii) an idealised MPC with perfect future knowledge, representing the theoretical best case performance.

6-1-1 Data usage

Table 6-1 summarizes how the available dataset is divided across the different stages of the study. The historical data from 2023-04-20 to 2024-04-20 is used to fit the ARMA models developed in Chapter 3, which provide the forecasts required for the predictive control framework. The subsequent period, from 2024-04-20 to 2025-04-20, is used for both training the RL-based control strategy and the evaluation of the derived methods. Specifically, 85% of this data is allocated for RL training to ensure robust policy learning under varying conditions, while the remaining 15% is reserved for performance evaluation and producing the final simulation results. The 15% is randomly sampled from the full year and represents roughly 50 days of simulations. This partitioning ensures that model training and testing are performed on distinct subsets, enabling an unbiased assessment of the proposed approach.

Data range	Amount	Usage
2023-04-20 to 2024-04-20	100%	Training ARMA models
2024-04-20 to 2025-04-20	85%	Training RL
2024-04-20 to 2025-04-20	15%	Producing results

Table 6-1: Data usage

6-1-2 Offer generation & analysis

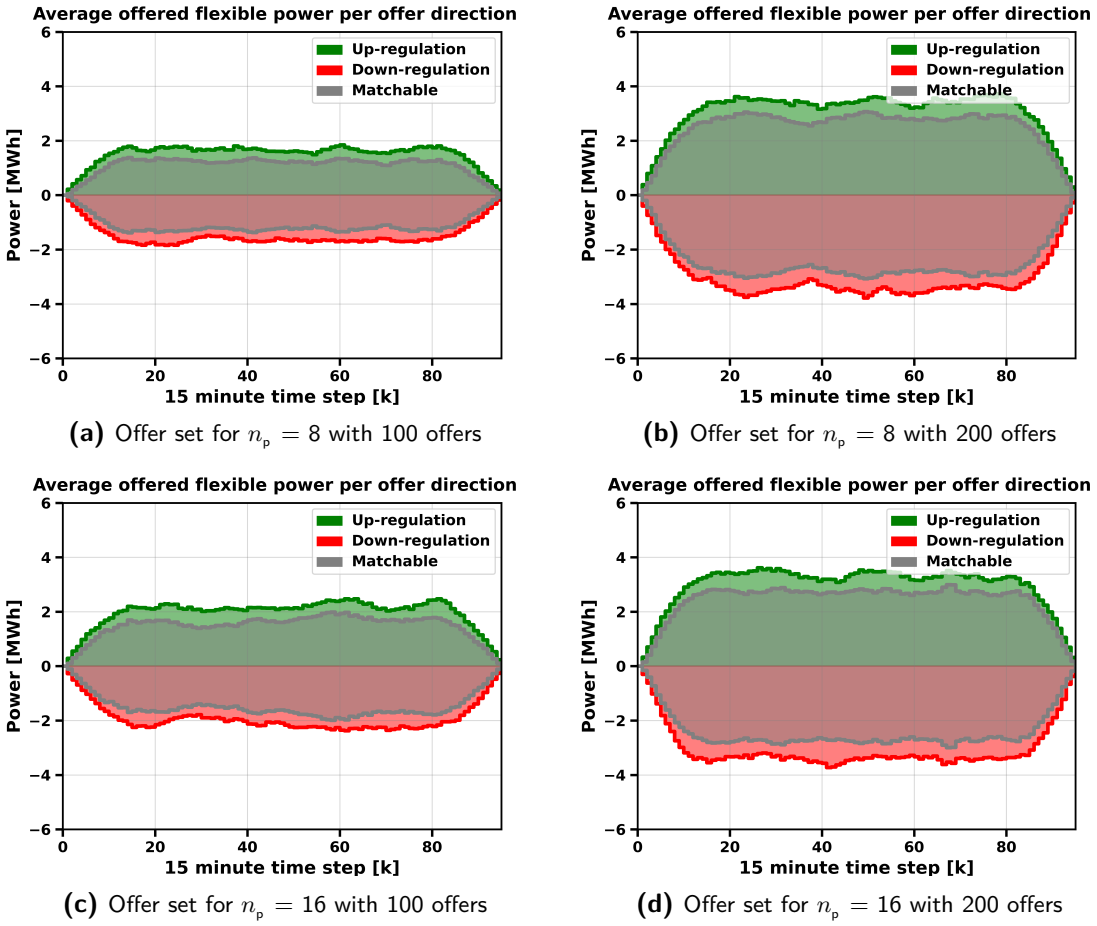
The offer set generation algorithm creates a collection of offers with randomized attributes to simulate market diversity. For each offer, all the variables defined in (4-18) and (4-14) are randomly allocated to simulate real-world market behaviour. The offer generation is described in Algorithm 2. In Figure 6-1 the average aggregated flexible power for the four offer sets is shown. The green and red shaded areas illustrate the available up- and down-regulation power, respectively, while the grey area denotes the matchable flexibility volume, i.e., the portion of power that could maximally be activated due to sufficient opposing offers. Figures 6-1a and 6-1b compare two offer set sizes for a prediction horizon of $n_{\mathrm{p}} = 8$. As

Algorithm 2: Offer Set Generation

for $i = \{1, \ldots, N_{offers}\}$ **do**
- $n \sim \{39, 40, 41, 42, 43, 44, 45, 46\}$
- $m^{(\mathrm{o})} \sim \{1, -1\}$
- **if** $m^{(o)} = 1$ **then**
  - $c^{(\mathrm{o})} \sim \mathcal{N}(180, 40)$
- **else**
  - $c^{(\mathrm{o})} \sim \mathcal{N}(-220, 50)$
- **end**
- $t^{\mathrm{start},(\mathrm{o})} \leftarrow UNIF(0, 96 - n_{\mathrm{p}})$
- $\ell^{\mathrm{min},(\mathrm{o})} \leftarrow UNIF(0, n_{\mathrm{p}})$
- $\ell^{\mathrm{max},(\mathrm{o})} \leftarrow UNIF(\ell^{\mathrm{min},(\mathrm{o})}, n_{\mathrm{p}})$
- $t^{\mathrm{stop},(\mathrm{o})} \leftarrow t^{\mathrm{start},(\mathrm{o})} + \ell^{\mathrm{max},(\mathrm{o})}$
- $\beta^{\mathrm{min},(\mathrm{o})} \leftarrow UNIF(0, 1)$
- **if** $FTO^{(o)}$ **then**
  - $P^{\mathrm{max},(\mathrm{o})} \leftarrow \mathcal{N}(m^{(\mathrm{o})}30, 10)$
  - $\mathcal{O} \leftarrow (n, t^{\mathrm{start},(\mathrm{o})}, t^{\mathrm{stop},(\mathrm{o})}, \ell^{\mathrm{min},(\mathrm{o})}, \ell^{\mathrm{max},(\mathrm{o})}, \beta^{\mathrm{min},(\mathrm{o})}, P^{\mathrm{max},(\mathrm{o})}, c^{(\mathrm{o})}, m^{(\mathrm{o})})$
- **else**
  - $P^{1,(\mathrm{o})} \leftarrow \mathcal{N}(m^{(\mathrm{o})}30, 10)$
  - $P^{2,(\mathrm{o})} \leftarrow \mathcal{N}(m^{(\mathrm{o})}30, 10)$
  - $P^{\mathrm{max},(\mathrm{o})} \leftarrow [P^{1,(\mathrm{o})} \ldots P^{1,(\mathrm{o})}\ P^{2,(\mathrm{o})} \ldots P^{2,(\mathrm{o})}]$
  - $\mathcal{O} \leftarrow (n, t^{\mathrm{start},(\mathrm{o})}, t^{\mathrm{stop},(\mathrm{o})}, \beta^{\mathrm{min},(\mathrm{o})}, P^{(\mathrm{o})}, c^{(\mathrm{o})}, m^{(\mathrm{o})})$
- **end**

end

expected, increasing the number of offers broadens both the up- and down-regulation ranges, resulting in greater overall flexibility and a larger matchable area. Figures 6-1c and 6-1d present the same comparison for a longer prediction horizon of $n_{\mathrm{p}} = 16$. Once again, the larger offer set provides higher flexibility across all time steps.



(a) Offer set for $n_p = 8$ with 100 offers

(b) Offer set for $n_p = 8$ with 200 offers

(c) Offer set for $n_p = 16$ with 100 offers

(d) Offer set for $n_p = 16$ with 200 offers

Figure 6-1: Aggregated flexible power for each of the offer sets averaged over all the different simulation days.

6-1-3 Case Study — 2024-08-20

In this section, a single day (2024-08-20) is analysed in detail to illustrate the behaviour and performance of the different control approaches. All methods are compared using the offer set corresponding to $n_p = 16$ and 200 available flexibility offers. The results are regrouped into three main figures to provide a concise overview of constraint violations, control actions, and aggregated performance metrics.

Figure 6-2 presents a comparison heatmap of the constraint violations across all transmission lines for each control strategy. The horizontal axis represents time, where each time step corresponds to a 15-minute interval, and the vertical axis lists the transmission lines. The colour intensity indicates the magnitude of the constraint violations, with darker shades of red representing higher levels of congestion.

The top panel shows the uncontrolled case, where one transmission line experiences persistent congestion while others become overloaded during the afternoon hours. The second panel displays the results of the Algorithmic Greedy Matching approach, which does not actively

deploy flexibility, resulting in similar congestion patterns to the uncontrolled case.

For the model-based controllers, the MPC method exhibits substantial activation of flexibility, leading to a clear reduction in the number and intensity of congested periods. The MPC-PP further improves congestion mitigation, confirming that forecast errors contribute to the residual violations seen in the standard MPC. Finally, the probabilistic controllers, CC-MPC and CC-MPC-RL, achieve the lowest overall congestion levels. These methods exhibit more preventive activation behaviour, mitigating minor overloads and demonstrating the benefits of probabilistic constraint handling in terms of robustness and reliability. Figure 6-3 shows

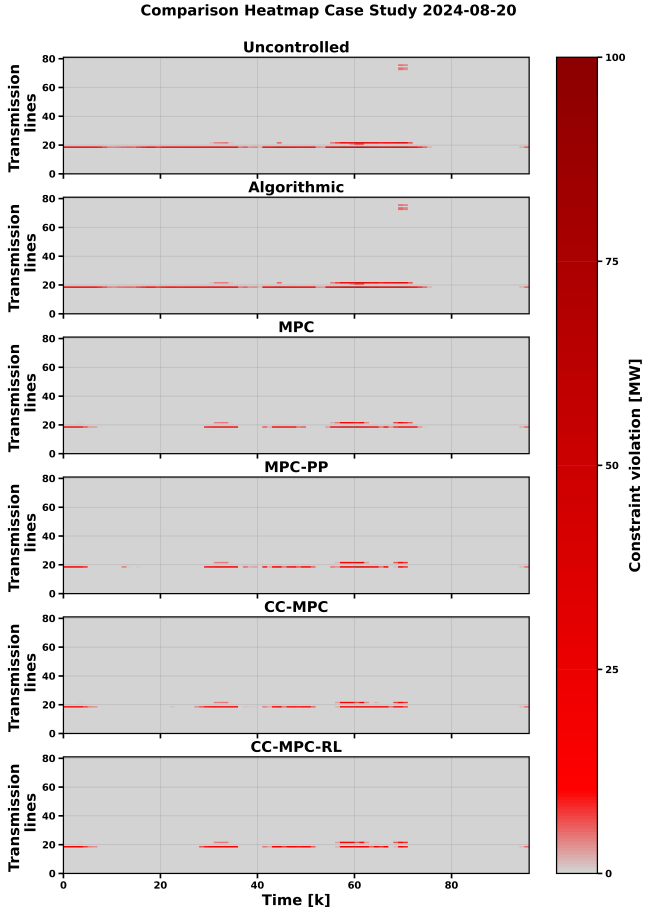


Figure 6-2: Comparison heatmap of constraint violations for all control strategies on 2024-08-20.

the corresponding control actions for the same day. Each subplot illustrates the up- and down-regulation activations of the respective control strategy, along with the total flexibility available in the system. The shaded grey area represents the total offered flexibility, while the green and red lines denote the activated up- and down-regulation, respectively.

The Algorithmic method shows no activations, consistent with the congestion observed in the heatmap. The MPC-based controllers activate flexibility throughout the day in response

to congestion events, with the MPC-PP showing slightly smoother activations due to perfect knowledge of future disturbances. The CC-MPC and CC-MPC-RL approaches display broader and more frequent activations, even in periods without visible congestion, reflecting a more conservative, preventive control strategy. While this increases operational cost, it significantly improves congestion prevention under uncertainty. The aggregated results for

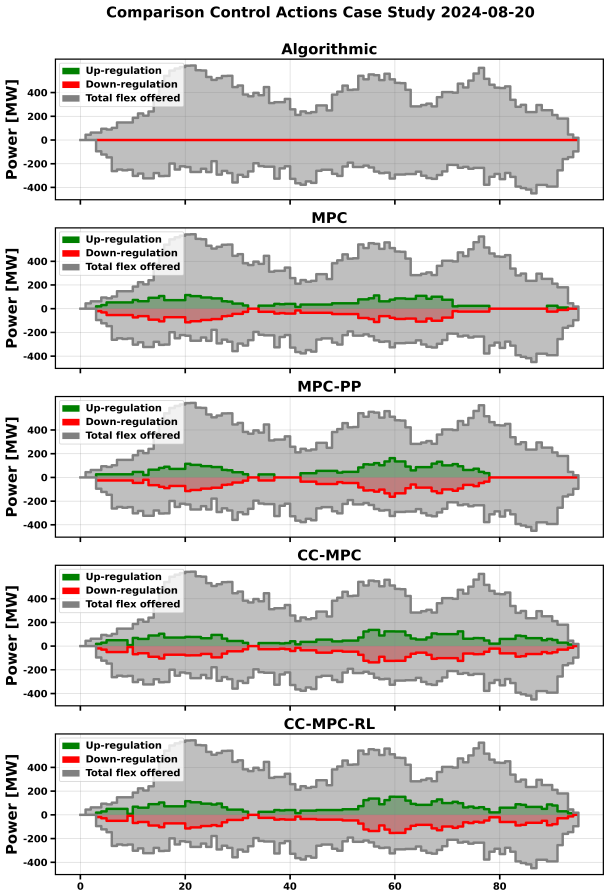


Figure 6-3: Comparison of control actions for all control strategies on 2024-08-20.

the case study are presented in Figure 6-4. The top panel shows the cumulative constraint violations, while the bottom panel presents the corresponding cumulative operational costs.

The uncontrolled and Algorithmic cases result in the highest cumulative violations, confirming the inability of simple matching or absence of control to alleviate congestion. The MPC-based methods achieve a substantial reduction in violations, with the perfect-prediction variant MPC-PP performing slightly better due to the absence of forecast uncertainty. The probabilistic controllers CC-MPC and CC-MPC-RL achieve the lowest cumulative violations overall, demonstrating that accounting for uncertainty explicitly improves reliability and robustness.

In terms of cumulative cost, the Algorithmic strategy yields the lowest value, as no flexibility activations occur. The MPC-based controllers incur higher costs due to increased flexibility usage, with the probabilistic methods being the most expensive. However, this higher cost corresponds to improved system security and reduced constraint violations. Notably, the CC-MPC-based methods achieve similar congestion mitigation performance to the perfect-prediction case while using imperfect forecasts, highlighting their robustness and practical applicability.

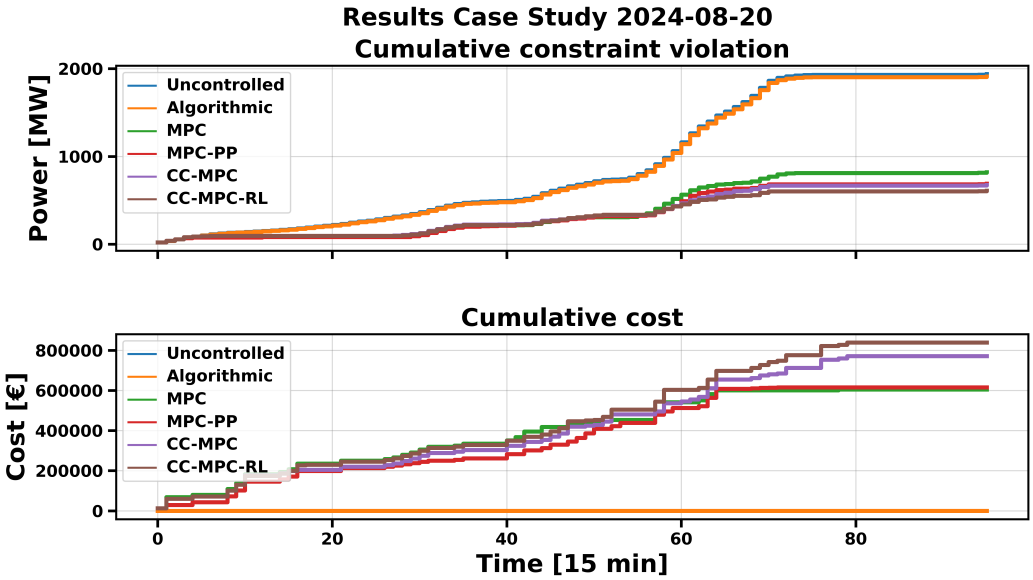


Figure 6-4: Cumulative constraint violation (top) and cumulative cost (bottom) for all control strategies during the 2024-08-20 case study.

6-2 Impact of Prediction Quality on Model Performance

As discussed in Chapter 3, the statistical models used for prediction might exhibit limited accuracy. To evaluate the impact of these prediction errors, the MPC-based approach is compared to MPC-PP, which uses perfect predictions, for different prediction horizons and number of offers. The results for both configurations, MPC and MPC-PP, are presented in Tables 6-2 and 6-3, respectively. Comparing Tables 6-2 and 6-3, the results confirm that prediction accuracy has a measurable impact on the performance of the nominal MPC. When perfect predictions are available, both the mean total violation and violation count are consistently lower, indicating that forecast uncertainty contributes to residual constraint violations. This effect becomes more pronounced for larger offer sets, where the controller has more flexibility to act on accurate forecasts. However, the differences in overall performance remain moderate, suggesting that even with imperfect predictions, the nominal MPC can effectively mitigate congestion. In other words, while forecast errors do reduce optimality, their impact does not critically undermine the controller's ability to operate the system efficiently.

MPC	$n_p = 8$ $n_{of} = 100$	$n_p = 8$ $n_{of} = 200$	$n_p = 16$ $n_{of} = 100$	$n_p = 16$ $n_{of} = 200$
Total violation (mean) [%]	-3.91 %	-6.57 %	-35.81 %	-56.18 %
Total violation (variance) [%]	5.97 %	7.74 %	25.0 %	26.78 %
Violation count (mean) [%]	-2.6 %	-4.68 %	-29.22 %	-46.65 %
Violation count (variance) [%]	7.42 %	9.58 %	22.91 %	27.55 %
Cost (mean) [€]	16692.68 €	22527.62 €	302572.72 €	538241.63 €
Cost (variance) [€]	11060.39 €	12117.94 €	219901.91 €	241400.13 €

Table 6-2: Summary of MPC performance metrics across 50 simulations for varying prediction horizons (n_p) and offer set sizes (n_{of}), relative to the uncontrolled case

MPC-PP	$n_p = 8$ $n_{of} = 100$	$n_p = 8$ $n_{of} = 200$	$n_p = 16$ $n_{of} = 100$	$n_p = 16$ $n_{of} = 200$
Total violation (mean) [%]	-4.54 %	-7.43 %	-41.49 %	-60.32 %
Total violation (variance) [%]	7.78 %	11.49 %	27.46 %	22.55 %
Violation count (mean) [%]	-1.94 %	-5.0 %	-32.19 %	-48.8 %
Violation count (variance) [%]	6.25 %	10.77 %	25.65 %	21.13 %
Cost (mean) [€]	14042.17 €	22273.09 €	313929.35 €	562637.84 €
Cost (variance) [€]	9827.81 €	14293.88 €	230628.23 €	258304.89 €

Table 6-3: Summary of MPC-PP performance metrics across 50 simulations for varying prediction horizons (n_p) and offer set sizes (n_{of}), relative to the uncontrolled case

6-3 Impact of risk parameter

One of the main advantages of CC-MPC is the ability to explicitly trade robustness for performance by adjusting the risk parameter α. Table 6-4 illustrates how varying α affects the mean and variance of violations and cost. As expected, a lower risk parameter ($\alpha = 0.75$) leads to a slightly lower cost, but at the cost of slightly increased constraint violation. The lower risk parameter ($\alpha = 0.75$) results in fewer control actions, leading to a slight reduction in operational costs but at the expense of increased constraint violations. In principle, relaxing the chance constraint (i.e., allowing a lower probability of constraint satisfaction) should enable the controller to operate more efficiently by prioritising performance over strict robustness. However, in this case, many of the congestions are structural and persist throughout the operating horizon. As a result, even when the controller is permitted to take on a higher level of risk, the underlying violations remain largely unavoidable. Consequently, the expected performance gain from reducing the risk parameter is limited, since the system cannot meaningfully exploit the additional flexibility provided by a lower α.

6-4 Impact of RL-Enhancement

The RL-enhanced variant CC-MPC-RL is introduced to automatically tune the variance of the uncertainty used in the prediction models based on observed closed-loop performance. However, in this work the RL training process was not fully stable: the learning curves exhibited oscillatory behaviour and no clear convergence to a single policy, as discussed in

CC-MPC	α	$n_p = 8$ $n_{of} = 100$	$n_p = 8$ $n_{of} = 200$	$n_p = 16$ $n_{of} = 100$	$n_p = 16$ $n_{of} = 200$
Total violation (mean) [%]	0.75	-3.03 %	-6.71 %	-39.38 %	-59.2 %
	0.9	-3.93 %	-7.13 %	-43.77 %	-59.44 %
Total violation (variance) [%]	0.75	7.93 %	11.24 %	25.52 %	24.82 %
	0.9	7.56 %	11.23 %	27.51 %	24.14 %
Violation count (mean) [%]	0.75	-1.47 %	-5.06 %	-32.27 %	-50.77 %
	0.9	-1.88 %	-5.74 %	-35.92 %	-50.85 %
Violation count (variance) [%]	0.75	6.38 %	12.87 %	22.9 %	23.43 %
	0.9	6.02 %	12.64 %	25.61 %	22.51 %
Cost (mean) [€]	0.75	12567.14 €	25642.35 €	355028.02 €	653568.64 €
	0.9	12258.20 €	25623.13 €	371277.57 €	690521.43 €
Cost (variance) [€]	0.75	11445.64 €	16194.16 €	237708.41 €	255827.52 €
	0.9	10249.82 €	17058.65 €	230599.56 €	238974.97 €

Table 6-4: Summary of CC-MPC performance metrics across 50 simulations for varying prediction horizons (n_p) and offer set sizes (n_{of}), relative to the uncontrolled case

Subsection 5-4. Table 6-5 summarizes the results for CC-MPC-RL. For the short prediction horizon $n_p = 8$ on which the RL-agent was trained, the proposed CC-MPC-RL achieves slightly lower mean violations than CC-MPC. Indicating that even though the RL training did not converge it still learned a policy capable of further reducing the constraint violation. For the configurations with a longer horizon ($n_p = 16$), the RL policy does not improve results, indicating that the learned policy is suitable to compensate and improve performance when a short horizon is chosen. Overall, the results show that RL can provide improvements, even if the full potential of the RL-enhanced design is not reached.

CC-MPC-RL	$n_p = 8$ $n_{of} = 100$	$n_p = 8$ $n_{of} = 200$	$n_p = 16$ $n_{of} = 100$	$n_p = 16$ $n_{of} = 200$
Total violation (mean) [%]	-4.2 %	-7.43 %	-41.88 %	-58.59 %
Total violation (variance) [%]	7.57 %	11.2 %	27.43 %	24.24 %
Violation count (mean) [%]	-2.2 %	-5.81 %	-35.82 %	-50.98 %
Violation count (variance) [%]	6.01 %	12.6 %	26.53 %	24.49 %
Cost (mean) [€]	10808.43 €	27051.14 €	387401.87 €	711028.66 €
Cost (variance) [€]	9913.18 €	18838.99 €	225910.49 €	222442.33 €

Table 6-5: Summary of CC-MPC-RL performance metrics across 50 simulations for varying prediction horizons (n_p) and offer set sizes (n_{of}), relative to the uncontrolled case

6-5 Comparison between methods

The previously shown results are summarized in Tables 6-6–6-9, where all methods are compared across the four configurations of prediction horizon ($n_p \in \{8, 16\}$) and offer set size ($n_{of} \in \{100, 200\}$). From this comparison, three consistent observations emerge.

Across all configurations, the Greedy Matching baseline performs the worst in terms of constraint violations. Among the predictive approaches, MPC-PP achieves the lowest mean

violation, benefiting from perfect foresight. The chance-constrained methods (CC-MPC and CC-MPC-RL) achieve nearly identical levels of constraint satisfaction while still relying on the same imperfect prediction models as the nominal MPC. This improvement, however, comes at a substantially higher cost, as the chance-constrained controllers activate flexibility pre-emptively in anticipation of possible violations—taking actions that the MPC-PP, with perfect future knowledge, would recognise as unnecessary.

Costs increase whenever violations are reduced by activating flexibility. The Greedy Matching approach is therefore cheapest but unacceptable from a risk/safety of operation perspective. Among the predictive methods, cost differences are closely related to the degree of violation reduction that is achieved, with MPC-PP representing an exception by demonstrating that accurate forecasting improves both constraint satisfaction and economic efficiency.

Moving from $n_{\mathrm{p}} = 8$ to $n_{\mathrm{p}} = 16$ substantially decreases violations for all predictive methods (e.g., MPC mean violation from -6.57% to -56.18% when $n_{\mathrm{of}} = 200$), showing that longer prediction horizons and richer offer sets enable more effective congestion mitigation. This highlights the importance of appropriate market order design to ensure that flexibility offers can be efficiently matched and utilised.

$n_{\mathrm{p}} = 8$ & $n_{\mathrm{of}} = 100$	Greedy Matching	MPC	CC-MPC	CC-MPC-RL	MPC-PP
Total violation (mean) [%]	-0.26 %	-3.91 %	-3.93 %	-4.2 %	-4.54 %
Total violation (variance) [%]	0 %	5.97 %	7.56 %	7.57 %	7.78 %
Violation count (mean) [%]	0 %	-2.6 %	-1.88 %	-2.2 %	-1.94 %
Violation count (variance) [%]	0 %	7.42 %	6.02 %	6.01 %	6.25 %
Cost (mean) [€]	1501.40 €	16692.68 €	12258.2 €	10808.43 €	14042.17 €
Cost (variance) [€]	0 €	11060.39 €	10249.82 €	9913.18 €	9827.81 €

Table 6-6: Performance comparison of control approaches with respect to the uncontrolled case for offer set with $n_{\mathrm{p}} = 8$ and $n_{\mathrm{of}} = 100$)

$n_{\mathrm{p}} = 8$ & $n_{\mathrm{of}} = 200$	Greedy Matching	MPC	CC-MPC	CC-MPC-RL	MPC-PP
Total violation (mean) [%]	-0.27 %	-6.57 %	-7.13 %	-7.43 %	-7.43 %
Total violation (variance) [%]	1.71 %	7.74 %	11.23 %	11.2 %	11.49 %
Violation count (mean) [%]	-0.32 %	-4.68 %	-5.74 %	-5.81 %	-5.0 %
Violation count (variance) [%]	0.77 %	9.58 %	12.64 %	12.6 %	10.77 %
Cost (mean) [€]	1944.92 €	22527.62 €	25623.13 €	27051.14 €	22273.09 €
Cost (variance) [€]	1181.43 €	12117.94 €	17058.65 €	18838.99 €	14293.88 €

Table 6-7: Performance comparison of control approaches with respect to the uncontrolled case for offer set with $n_{\mathrm{p}} = 8$ and $n_{\mathrm{of}} = 200$)

$n_{\mathrm{p}} = 16$ & $n_{\mathrm{of}} = 100$	Greedy Matching	MPC	CC-MPC	CC-MPC-RL	MPC-PP
Total violation (mean) [%]	-8.83 %	-35.81 %	-43.77%	-41.88 %	-41.49 %
Total violation (variance) [%]	11.89 %	25.0 %	27.51 %	27.43 %	27.46 %
Violation count (mean) [%]	-7.76 %	-29.22 %	-35.92 %	-35.82 %	-32.19 %
Violation count (variance) [%]	9.94 %	22.91 %	25.61 %	26.53 %	25.65 %
Cost (mean) [€]	102719.91 €	302572.72 €	371277.57 €	387401.87 €	313929.35 €
Cost (variance) [€]	59598.27 €	219901.91 €	230599.56 €	225910.49 €	230628.23 €

Table 6-8: Performance comparison of control approaches with respect to the uncontrolled case for offer set with $n_{\mathrm{p}} = 16$ and $n_{\mathrm{of}} = 100$)

$n_p = 16$ & $n_{of} = 200$	Greedy Matching	MPC	CC-MPC	CC-MPC-RL	MPC-PP
Total violation (mean) [%]	-10.19 %	-56.18 %	-59.44 %	-58.59 %	-60.32 %
Total violation (variance) [%]	14.26 %	26.78 %	24.14 %	24.24 %	22.55 %
Violation count (mean) [%]	-5.53 %	-46.65 %	-50.85 %	-50.98 %	-48.8 %
Violation count (variance) [%]	6.46 %	27.55 %	22.51 %	24.49 %	21.13 %
Cost (mean) [€]	146396.78 €	538241.63 €	690521.43 €	711028.66 €	562637.84 €
Cost (variance) [€]	80776.80 €	241400.13 €	238974.97 €	222442.33 €	258304.89 €

Table 6-9: Performance comparison of control approaches with respect to the uncontrolled case for offer set with $n_p = 16$ and $n_{of} = 200$)

6-6 Summary

This chapter presented and compared the performance of the proposed congestion management methods across a range of modelling and operational conditions. The analysis considered variations in prediction horizon, risk parameter, prediction accuracy, and offer quantity, supported by a structured simulation set-up using both synthetic market offers and real historical data. A detailed day-ahead case study (2024-08-20) was used to illustrate the differences between approaches, showing that uncontrolled operation and Greedy Matching fail to mitigate congestion effectively, while all model-based predictive methods achieve substantial reductions in line overloads.

The MPC-based controllers successfully coordinated up- and down-regulation actions to maintain system balance, even when relying on imperfect SARIMA-based forecasts. Despite relying on approximate uncertainty models, the chance-constrained approach achieves performance close to that of MPC-PP, which has access to perfect future information, albeit at a substantially higher cost due to its added conservatism. The CC-MPC method, which introduces probabilistic constraints, achieved even better constraint satisfaction and nearly matched the performance of MPC-PP while using the same imperfect forecasts. Adjusting the risk parameter α allowed explicit control over the robustness–performance trade-off, though the structural nature of congestion limited its overall impact.

The analysis of offer availability showed that increasing the number of offers and extending the prediction horizon both enhance the system's controllability by enlarging the pool of matchable flexibility. However, these gains come at the cost of increased activations and computational effort, with diminishing returns once sufficient flexibility is available. The RL-enhanced CC-MPC approach (CC-MPC-RL) demonstrated potential for automated adaptation but suffered from unstable training and poor generalisation to unseen configurations. While it outperformed CC-MPC for the configuration it was trained on, improvements were inconsistent across other cases.

Overall, the results confirm that predictive and probabilistic control strategies are highly effective for managing structural network congestion. Even with imperfect forecasts, the proposed approaches significantly reduce violations compared to static or greedy alternatives.

Chapter 7

Conclusion, Discussion and Future work

7-1 Conclusion

The objective of this work has been captured by the following research question:

> *How can an MPC-based CM strategy, enhanced with RL, be implemented to manage congestion under uncertainty for the Dutch transmission grid using real-world data?*

This thesis addressed these questions through a structured methodology combining system modelling, control theory, data analysis, and simulation-based validation. First, the Dutch transmission grid was represented within an MPC framework, allowing the CM problem to be expressed as a dynamic optimisation problem. Then, the MPC problem was extended to incorporate the market rules of CM in the Netherlands, allowing the interaction between technical control and market-clearing mechanisms to be captured. Real-world data from the Dutch grid have been utilized to form a statistical model, incorporating realistic uncertainty in the network. The inclusion of RL enhanced the CC-MPC framework by enabling adaptive tuning of control parameters and improved decision-making under uncertain and time-varying conditions.

The simulation and case study results demonstrated that the combined CC-MPC-RL approach can effectively anticipate congestion events, coordinate flexibility activation, and reduce congestion events compared to other methods. The integration of statistical uncertainty models and a stochastic CC-MPC formulation improved robustness under uncertainty, while the RL component allowed the controller to improve results. Overall, the findings confirm the feasibility and potential of an CC-MPC-based CM strategy augmented with RL to enhance grid operation in the Dutch context.

The control architecture integrates a CC-MPC scheme with an adaptive RL component for online uncertainty tuning. The results indicate that the CC-MPC approach consistently

outperforms the deterministic MPC and greedy matching strategies in terms of constraint satisfaction. This confirms that incorporating probabilistic forecasts can effectively balance performance and security. Moreover, the inclusion of the RL layer enables adaptive variance estimation of the stochastic model, thereby enhancing robustness under varying system conditions.

7-2 Discussion

The proposed framework successfully bridges the gap between technical control methodologies and market-based CM. While the results demonstrate improvements in both operational and economic performance, several aspects merit further reflection and refinement.

The CC-MPC and its RL-enhanced variant require the repeated solution of multiple stochastic scenarios, which increases computation time and may challenge real-time implementation for large-scale networks. This trade-off between control accuracy and computational tractability is central to future deployment considerations.

The control framework relies on a linearised transmission network model derived from the AC power flow equations. While this linearisation enables efficient optimisation, it is valid primarily for high-voltage transmission networks with small voltage angle differences. Because the proposed platform is also intended to address congestion in medium- and low-voltage networks, this simplification constitutes a significant limitation. Incorporating non-linear AC formulations or reduced-order non-linear models would improve accuracy and extend applicability across voltage levels; however, at significant computational cost.

The GOPACS market representation adopted in this study presumes greater flexibility availability than is presently observed in the real system. In practice, both the quantity and scale of market offers are considerably lower, which restricts the impact the proposed method can have.

The statistical disturbance model developed using data from EDSN formed the basis for uncertainty representation within the proposed CC-MPC framework. Although the implemented SARIMA models captured general production patterns, forecasting accuracy varied significantly between regions. The results demonstrate that model accuracy directly influences control performance, as improved forecasts lead to more effective congestion mitigation. This underlines the need for more advanced and adaptive forecasting techniques.

Residual analysis revealed mild heteroscedasticity, indicating that forecast uncertainty is time-varying and strongly influenced by weather-driven variability in renewable generation. Future work should therefore explore the integration of volatility models such as generalized autoregressive conditional heteroscedastic models, which explicitly account for time-varying variance and could provide a more accurate representation of stochastic behaviour in renewable energy forecasts (e.g., ensemble models, volatility extensions to SARIMA, and or the incorporation of exogenous variables, or deep learning models).

In this work only two offer structures were investigated; however, in practical CM, more complex offer types—such as exclusive, block, or linked offers—are common and might be needed to attract the necessary amounts of flexibility. Expanding the market model to accommodate these advanced offer structures could facilitate more users.

Additionally, the performance of the RL component warrants further investigation, particularly with regards to training convergence.

Overall, the findings confirm that combining probabilistic forecasting, CC-MPC, and RL provides a promising and powerful approach to CM under uncertainty. Extending this framework to non-linear AC formulations, integrating advanced forecasting techniques, incorporating richer market order types, and expanding its scope to include additional parts of the power grid represent promising directions for future research and development.

7-3 Contributions and Future Work

7-3-1 Contributions

This thesis contributes to the ongoing research on market-based CM in by developing a novel CC-MPC approach that bridges the gap between advanced theoretical control techniques in literature to real-world applicable control strategies. First, a representative market model of the Dutch congestion management framework is developed and integrated into a dynamic CC-MPC-based optimisation framework.

Second, it presents an integrated control framework that combines CC-MPC with RL to enhance adaptability under stochastic operating conditions. This hybrid design enables the controller to dynamically tune the uncertainty representation in response to observed performance, improving constraint satisfaction compared to deterministic MPC and CC-MPC methods.

Second, the thesis introduces a comprehensive data-driven approach to uncertainty quantification for real-world CM applications. By leveraging time-series data from EDSN, statistical models were constructed to represent renewable generation variability and used directly within the CC-MPC optimisation. This integration of real-world data into a predictive control setting demonstrates a practical pathway to incorporating uncertainty directly in the decision making process.

Third, the thesis validates the proposed methods through a case study based on the Dutch high-voltage transmission grid. The study compares multiple control strategies; greedy matching, deterministic MPC, CC-MPC, and CC-MPC–RL, and demonstrates the superior performance of the proposed approach in congestion mitigation under uncertainty.

7-3-2 Future Work

Building on the findings of this work, several avenues for future research can be identified to strengthen and extend the proposed framework. First, enhancing uncertainty modelling remains a key priority. Since the forecasting quality of the SARIMA-based models was shown to influence control performance, future research should focus on integrating more advanced uncertainty representations, such as ensemble models, volatility extensions to SARIMA, the inclusion of exogenous variables, or deep learning-based approaches. In addition, the applicability of the controller could be improved by addressing the limitations of the linearised grid model. Extending the framework to a non-linear AC formulation would allow the method to

capture lower grid levels, thereby making it relevant for both transmission and distribution networks. Finally, expanding the market representation to include complex offer types, such as linked, block, and exclusive offers, would enable a more complete assessment of flexibility trading mechanisms in large-scale CM. As market design strongly influences how flexibility is activated and valued, these extensions are essential for evaluating the framework under realistic market conditions. Overall, future work should refine the proposed method to enhance its applicability in real-world systems.

Appendix A

Derivation of the Linearised Model

Starting from the non-linear AC power flow equations (2-3) and the nodal power balance equations (2-2), a linear, time-varying dynamic network model is derived. The model is expressed as

$$\boldsymbol{x}(k+1) = S(k)\boldsymbol{x}(k) + T(k)\boldsymbol{u}(k) + T(k)\boldsymbol{w}(k), \tag{A-1}$$

where $S(k)$ and $T(k)$ are time-varying system matrices, $\boldsymbol{x}(k)$ is the state vector, $\boldsymbol{u}(k)$ collects controllable power changes, and $\boldsymbol{w}(k)$ collects uncontrollable power changes.

Let $\mathscr{N}$ be the set of nodes and $\mathscr{E}$ the set of directed transmission lines, with $|\mathscr{N}|$ the number of nodes and $|\mathscr{E}|$ the number of lines. The stacked state, input, and disturbance vectors are defined as

$$\boldsymbol{x}(k) = \begin{bmatrix} P^{(1)}(k) \\ \vdots \\ P^{(|\mathscr{N}|)}(k) \\ Q^{(1)}(k) \\ \vdots \\ Q^{(|\mathscr{N}|)}(k) \\ P^{(1,2)}(k) \\ \vdots \\ P^{(|\mathscr{E}|)}(k) \\ Q^{(1,2)}(k) \\ \vdots \\ Q^{(|\mathscr{E}|)}(k) \end{bmatrix}, \quad \boldsymbol{u}(k) = \begin{bmatrix} \Delta P_u^{(1)}(k) \\ \vdots \\ \Delta P_u^{(n_u)}(k) \\ \Delta Q_u^{(1)}(k) \\ \vdots \\ \Delta Q_u^{(n_u)}(k) \end{bmatrix}, \quad \boldsymbol{w}(k) = \begin{bmatrix} \Delta P_w^{(1)}(k) \\ \vdots \\ \Delta P_w^{(n_w)}(k) \\ \Delta Q_w^{(1)}(k) \\ \vdots \\ \Delta Q_w^{(n_w)}(k) \end{bmatrix}, \tag{A-2}$$

where $P^{(n)}(k)$ and $Q^{(n)}(k)$ are respectively the active and reactive nodal powers at node $n \in \mathscr{N}$, and $P^{(n,m)}(k)$, $Q^{(n,m)}(k)$ are the active and reactive power flows on line $(n,m) \in \mathscr{E}$. For each node $n \in \mathscr{N}$ define,

$$\begin{aligned} P^{(n)}(k+1) &= P^{(n)}(k) + \Delta P^{(n)}(k), \\ Q^{(n)}(k+1) &= Q^{(n)}(k) + \Delta Q^{(n)}(k). \end{aligned} \tag{A-3}$$

The nodal power changes are decomposed into controllable and uncontrollable components:

$$\begin{aligned}\Delta P^{(n)}(k) &= \Delta P_u^{(n)}(k) + \Delta P_w^{(n)}(k),\\ \Delta Q^{(n)}(k) &= \Delta Q_u^{(n)}(k) + \Delta Q_w^{(n)}(k),\end{aligned} \tag{A-4}$$

where $\Delta P_u^{(n)}(k)$ and $\Delta Q_u^{(n)}(k)$ are controllable active and reactive power changes (for example, from generators), and $\Delta P_w^{(n)}(k)$ and $\Delta Q_w^{(n)}(k)$ are uncontrollable changes (for example, from loads or renewable generation).

The line power flows between nodes n and m are given by the non-linear AC power flow expressions in (2-3)

$$\begin{aligned}P^{(n,m)}(k) &= f_{P^{(n,m)}}(a(k)),\\ Q^{(n,m)}(k) &= f_{Q^{(n,m)}}(a(k)),\end{aligned} \tag{A-5}$$

where

$$a(k) = \{v^{(n)}(k), v^{(m)}(k), \theta^{(n)}(k), \theta^{(m)}(k)\} \tag{A-6}$$

collects the voltage magnitudes $v^{(\cdot)}(k)$ and phase angles $\theta^{(\cdot)}(k)$ at the line end nodes. Define the corresponding changes

$$\Delta a(k) = a(k+1) - a(k). \tag{A-7}$$

A first-order Taylor expansion of (A-5) around the operating point $a(k)$ yields

$$P^{(n,m)}(k+1) \approx P^{(n,m)}(k) + \left.\frac{\partial f_{P^{(n,m)}}}{\partial a}\right|_{a(k)} \Delta a(k), \tag{A-8a}$$

$$Q^{(n,m)}(k+1) \approx Q^{(n,m)}(k) + \left.\frac{\partial f_{Q^{(n,m)}}}{\partial a}\right|_{a(k)} \Delta a(k), \tag{A-8b}$$

which implies

$$\begin{aligned}\Delta P^{(n,m)}(k) &= \left.\frac{\partial f_{P^{(n,m)}}}{\partial a}\right|_{a(k)} \Delta a(k),\\ \Delta Q^{(n,m)}(k) &= \left.\frac{\partial f_{Q^{(n,m)}}}{\partial a}\right|_{a(k)} \Delta a(k).\end{aligned} \tag{A-9}$$

Stacking (A-9) for all lines $(n,m) \in \mathscr{E}$ yields

$$\begin{bmatrix}\Delta \boldsymbol{P}^{(n,m)}(k)\\ \Delta \boldsymbol{Q}^{(n,m)}(k)\end{bmatrix} = \begin{bmatrix}H_{P\theta}(k) & H_{Pv}(k)\\ H_{Q\theta}(k) & H_{Qv}(k)\end{bmatrix} \begin{bmatrix}\Delta \boldsymbol{\theta}^{(n)}(k)\\ \Delta \boldsymbol{v}^{(n)}(k)\end{bmatrix}, \tag{A-10}$$

where

$$\Delta \boldsymbol{P}^{(n,m)}(k) = \begin{bmatrix}\Delta P^{(1,2)}(k)\\ \vdots\\ \Delta P^{(|\mathscr{E}|)}(k)\end{bmatrix}, \quad \Delta \boldsymbol{Q}^{(n,m)}(k) = \begin{bmatrix}\Delta Q^{(1,2)}(k)\\ \vdots\\ \Delta Q^{(|\mathscr{E}|)}(k)\end{bmatrix},$$

$$\Delta \boldsymbol{\theta}^{(n)}(k) = \begin{bmatrix}\Delta \theta^{(1)}(k)\\ \vdots\\ \Delta \theta^{(|\mathscr{N}|)}(k)\end{bmatrix}, \quad \Delta \boldsymbol{v}^{(n)}(k) = \begin{bmatrix}\Delta v^{(1)}(k)\\ \vdots\\ \Delta v^{(|\mathscr{N}|)}(k)\end{bmatrix},$$

and the matrices $H_{P\theta}(k)$, $H_{Pv}(k)$, $H_{Q\theta}(k)$, and $H_{Qv}(k)$ are the Jacobians of line active and reactive power flows with respect to the nodal voltage angles and magnitudes, evaluated at time k.

The nodal active and reactive powers are given by the non-linear power balance equations in (2-2)

$$\begin{aligned} P^{(n)}(k) &= f_{P^{(n)}}(b(k)), \\ Q^{(n)}(k) &= f_{Q^{(n)}}(b(k)), \end{aligned} \tag{A-11}$$

where

$$b(k) = \{v^{(1)}(k), \ldots, v^{(|\mathcal{N}|)}(k), \theta^{(1)}(k), \ldots, \theta^{(|\mathcal{N}|)}(k)\} \tag{A-12}$$

collects the voltage magnitudes and angles at all nodes. Denote the corresponding change

$$\Delta b(k) = b(k+1) - b(k). \tag{A-13}$$

A first-order Taylor expansion of the nodal powers yields, for each node n,

$$\Delta P^{(n)}(k) = \frac{\partial f_{P^{(n)}}}{\partial b}\bigg|_{b(k)} \Delta b(k), \tag{A-14a}$$

$$\Delta Q^{(n)}(k) = \frac{\partial f_{Q^{(n)}}}{\partial b}\bigg|_{b(k)} \Delta b(k). \tag{A-14b}$$

Stacking these for all nodes gives

$$\begin{bmatrix} \Delta \boldsymbol{P}^{(n)}(k) \\ \Delta \boldsymbol{Q}^{(n)}(k) \end{bmatrix} = \begin{bmatrix} J_{P\theta}(k) & J_{Pv}(k) \\ J_{Q\theta}(k) & J_{Qv}(k) \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{\theta}^{(n)}(k) \\ \Delta \boldsymbol{v}^{(n)}(k) \end{bmatrix}, \tag{A-15}$$

where

$$\Delta \boldsymbol{P}^{(n)}(k) = \begin{bmatrix} \Delta P^{(1)}(k) \\ \vdots \\ \Delta P^{(|\mathcal{N}|)}(k) \end{bmatrix}, \quad \Delta \boldsymbol{Q}^{(n)}(k) = \begin{bmatrix} \Delta Q^{(1)}(k) \\ \vdots \\ \Delta Q^{(|\mathcal{N}|)}(k) \end{bmatrix},$$

and the Jacobian submatrices are defined as

$$\begin{aligned} J_{P\theta}(k) &= \frac{\partial \boldsymbol{P}^{(|\mathcal{N}|)}(k)}{\partial \boldsymbol{\theta}^{(n)}(k)}, \quad J_{Pv}(k) = \frac{\partial \boldsymbol{P}^{(|\mathcal{N}|)}(k)}{\partial \boldsymbol{v}^{(n)}(k)}, \\ J_{Q\theta}(k) &= \frac{\partial \boldsymbol{Q}^{(|\mathcal{N}|)}(k)}{\partial \boldsymbol{\theta}^{(n)}(k)}, \quad J_{Qv}(k) = \frac{\partial \boldsymbol{Q}^{(|\mathcal{N}|)}(k)}{\partial \boldsymbol{v}^{(n)}(k)}. \end{aligned} \tag{A-16}$$

Assuming that the nodal Jacobian

$$J(k) := \begin{bmatrix} J_{P\theta}(k) & J_{Pv}(k) \\ J_{Q\theta}(k) & J_{Qv}(k) \end{bmatrix} \tag{A-17}$$

is nonsingular at the operating point, (A-15) can be inverted to express the voltage angle and magnitude changes in terms of the nodal power changes:

$$\begin{bmatrix} \Delta \boldsymbol{\theta}^{(n)}(k) \\ \Delta \boldsymbol{v}^{(n)}(k) \end{bmatrix} = J(k)^{-1} \begin{bmatrix} \Delta \boldsymbol{P}^{(n)}(k) \\ \Delta \boldsymbol{Q}^{(n)}(k) \end{bmatrix}. \tag{A-18}$$

Substituting (A-18) into the line-flow relation (A-10) gives

$$\begin{bmatrix}\Delta \boldsymbol{P}^{(n,m)}(k)\\ \Delta \boldsymbol{Q}^{(n,m)}(k)\end{bmatrix} = \begin{bmatrix}H_{P\theta}(k) & H_{Pv}(k)\\ H_{Q\theta}(k) & H_{Qv}(k)\end{bmatrix} J(k)^{-1} \begin{bmatrix}\Delta \boldsymbol{P}^{(n)}(k)\\ \Delta \boldsymbol{Q}^{(n)}(k)\end{bmatrix}. \tag{A-19}$$

Define the composite sensitivity matrix

$$G(k) := \begin{bmatrix}H_{P\theta}(k) & H_{Pv}(k)\\ H_{Q\theta}(k) & H_{Qv}(k)\end{bmatrix} J(k)^{-1} = \begin{bmatrix}G_{\mathrm{pp}}(k) & G_{\mathrm{pq}}(k)\\ G_{\mathrm{qp}}(k) & G_{\mathrm{qq}}(k)\end{bmatrix}, \tag{A-20}$$

where the block matrices $G_{\mathrm{pp}}(k)$, $G_{\mathrm{pq}}(k)$, $G_{\mathrm{qp}}(k)$, and $G_{\mathrm{qq}}(k)$ have dimensions $|\mathscr{E}| \times |\mathscr{N}|$ and collect the sensitivities from nodal active/reactive power changes to line active/reactive power changes. The relation between nodal and line power changes can then be written as

$$\begin{bmatrix}\Delta \boldsymbol{P}^{(n,m)}(k)\\ \Delta \boldsymbol{Q}^{(n,m)}(k)\end{bmatrix} = \begin{bmatrix}G_{\mathrm{pp}}(k) & G_{\mathrm{pq}}(k)\\ G_{\mathrm{qp}}(k) & G_{\mathrm{qq}}(k)\end{bmatrix} \begin{bmatrix}\Delta \boldsymbol{P}^{(n)}(k)\\ \Delta \boldsymbol{Q}^{(n)}(k)\end{bmatrix}. \tag{A-21}$$

Writing (A-21) component-wise, for each line $(n,m) \in \mathscr{E}$,

$$\Delta P^{(n,m)}(k) = \sum_{l \in \mathscr{N}} g_{\mathrm{pp}}^{(n,m),l}(k)\, \Delta P^{(l)}(k) + \sum_{l \in \mathscr{N}} g_{\mathrm{pq}}^{(n,m),l}(k)\, \Delta Q^{(l)}(k), \tag{A-22a}$$

$$\Delta Q^{(n,m)}(k) = \sum_{l \in \mathscr{N}} g_{\mathrm{qp}}^{(n,m),l}(k)\, \Delta P^{(l)}(k) + \sum_{l \in \mathscr{N}} g_{\mathrm{qq}}^{(n,m),l}(k)\, \Delta Q^{(l)}(k), \tag{A-22b}$$

where $g_{\mathrm{pp}}^{(n,m),l}(k)$ is the sensitivity of the active power flow on line (n,m) with respect to a change in active power injection at node l; $g_{\mathrm{pq}}^{(n,m),l}(k)$ is the sensitivity of the same active line flow with respect to a change in reactive power injection at node l; $g_{\mathrm{qp}}^{(n,m),l}(k)$ and $g_{\mathrm{qq}}^{(n,m),l}(k)$ are defined analogously for reactive line flows.

Stacking the coefficients in (A-22) for all lines yields the block matrices

$$\begin{aligned}
G_{\mathrm{pp}}(k) &= \begin{bmatrix} g_{\mathrm{pp}}^{(1,2),1}(k) & \dots & g_{\mathrm{pp}}^{(1,2),|\mathscr{N}|}(k) \\ \vdots & \ddots & \vdots \\ g_{\mathrm{pp}}^{(|\mathscr{E}|),1}(k) & \dots & g_{\mathrm{pp}}^{(|\mathscr{E}|),|\mathscr{N}|}(k) \end{bmatrix}, \\
G_{\mathrm{pq}}(k) &= \begin{bmatrix} g_{\mathrm{pq}}^{(1,2),1}(k) & \dots & g_{\mathrm{pq}}^{(1,2),|\mathscr{N}|}(k) \\ \vdots & \ddots & \vdots \\ g_{\mathrm{pq}}^{(|\mathscr{E}|),1}(k) & \dots & g_{\mathrm{pq}}^{(|\mathscr{E}|),|\mathscr{N}|}(k) \end{bmatrix}, \\
G_{\mathrm{qp}}(k) &= \begin{bmatrix} g_{\mathrm{qp}}^{(1,2),1}(k) & \dots & g_{\mathrm{qp}}^{(1,2),|\mathscr{N}|}(k) \\ \vdots & \ddots & \vdots \\ g_{\mathrm{qp}}^{(|\mathscr{E}|),1}(k) & \dots & g_{\mathrm{qp}}^{(|\mathscr{E}|),|\mathscr{N}|}(k) \end{bmatrix}, \\
G_{\mathrm{qq}}(k) &= \begin{bmatrix} g_{\mathrm{qq}}^{(1,2),1}(k) & \dots & g_{\mathrm{qq}}^{(1,2),|\mathscr{N}|}(k) \\ \vdots & \ddots & \vdots \\ g_{\mathrm{qq}}^{(|\mathscr{E}|),1}(k) & \dots & g_{\mathrm{qq}}^{(|\mathscr{E}|),|\mathscr{N}|}(k) \end{bmatrix}.
\end{aligned} \tag{A-23}$$

The matrices $G_{\mathrm{pp}}(k)$, $G_{\mathrm{pq}}(k)$, $G_{\mathrm{qp}}(k)$, and $G_{\mathrm{qq}}(k)$ therefore describe how changes in nodal active and reactive powers propagate to changes in active and reactive power flows on the transmission lines. These sensitivity matrices can then be used to assemble the time-varying matrices $S(k)$ and $T(k)$ in (A-1), yielding a linear, discrete-time dynamic model of the AC network that captures both controllable and uncontrollable power variations.

Appendix B

Paper style thesis

A Stochastic Learning-based Model Predictive Control Approach for Market-Based Congestion Management

1st J. van der Weerd
Delft Center for Systems and Control
TU Delft
Delft, the Netherlands

2nd F. Cordiano
Delft Center for Systems and Control
TU Delft
Delft, the Netherlands

3rd A. Riccardi
Delft Center for Systems and Control
TU Delft
Delft, the Netherlands

4th B.H.K. De Schutter
Delft Center for Systems and Control
TU Delft
Delft, the Netherlands

***Abstract*—The rapid growth of renewable energy sources (RES) and electrification of transport, heating, and industry are transforming the Dutch power grid. While crucial for climate goals, these trends introduce uncertainty and complicate network control. Existing Congestion Management (CM) approaches often overlook the stochastic nature of Renewable Energy Source (RES) generation, simplify network models, or ignore market constraints. This work formulates market-based CM for the Dutch grid as a Chance-Constrained Model Predictive Control (CC-MPC) problem. A linearised high-voltage network model is integrated with a mixed-integer formulation of market offers. RES uncertainty is represented via Seasonal AutoRegressive Integrated Moving-Average (SARIMA)-based forecasts. To favour the trade off between congestion cost and constraint satisfaction, a Reinforcement Learning approach is introduced, to actively tune a scaling parameters for the variance of the uncertainty. Then, the sampling-based approximation of the CC-MPC determines which flexibility offers need to be accepted. Using real data from Energie Data Services Nederland (EDSN) and Grid Operators Platform for AnCillary Services (GOPACS), results show that the Reinforcement Learning (RL)-enhanced CC-MPC improves constraint satisfaction, demonstrating superior results compared to other traditional CM methods.**

I. Introduction

The Paris Agreement set global commitments to limit the temperature rise by rapidly reducing greenhouse gas emissions [1]. RESs such as wind and solar are central to this effort, with governments pledging to triple renewable capacity by 2030 [2]. Simultaneously, electrification of transport, heating, and industry is accelerating electricity demand, driven by both climate and energy security goals [3]. Meeting this demand requires substantial grid investments, estimated at €70 billion annually in the European Union until 2050 [4].

This large-scale transformation challenges grid operators, as existing networks were designed for centralized, unidirectional power flows. Increasing RES penetration introduces intermittency, while reduced conventional generation limits system flexibility, amplifying congestion risks [5]. Such congestion can cause voltage instability, equipment overloading, and costly remedial actions, amounting to €4.26 billion in 2023 alone [6].

Efficient CM is thus vital to ensure secure, economical, and sustainable grid operation. Traditional methods are often overly simplistic, while advanced optimisation-based approaches proposed in the literature frequently neglect real-world limitations. To bridge this gap CC-MPC has emerged as a promising framework for handling complex system dynamics, operational constraints, and uncertainty. In this context, statistical models and RL techniques are used to enhance adaptability and performance by learning from evolving grid conditions.

This work makes three key contributions to the field of market-based CM:

- It introduces a novel hybrid CC-MPC–RL control framework that enables adaptive, uncertainty-aware CM. The reinforcement learning layer dynamically adjusts uncertainty parameters to improve robustness and constraint satisfaction.
- It presents a data-driven methodology for uncertainty quantification using real-world EDSN data. This approach directly embeds statistical forecasts into the predictive control problem, demonstrating a practical route for integrating real-world variability into model-based decision-making.
- It validates the proposed methods through case studies on the Dutch transmission grid, comparing the hybrid approach with greedy and deterministic baselines. The results show consistent improvements in congestion mitigation and robustness under uncertain operating conditions.

The remainder of this paper is organised as follows. Section II introduces the modelling framework, describing the grid topology, linearised network dynamics, and market representation. Section III outlines the data sources and stochastic forecasting

approach used to represent renewable generation uncertainty. Sections IV and V presents the proposed CC-MPC formulation and its reinforcement learning enhancement. Section VI details the case study and simulation results, demonstrating the effectiveness of the proposed approach. Finally, Section VII summarises the main findings and discusses potential directions for future work.

II. Modelling

This section presents the model of the high-voltage grid of the Netherlands (Section II-A), the linearised dynamical model of a high-voltage transmission network (Section II-B), and the market model derived from the Dutch CM platform GOPACS (Section II-C), which together form the basis of the proposed CC-MPC framework.

A. Grid model

The Dutch high-voltage grid is represented as a undirected mathematical graph, based on the map made by Tennet [7], consisting of nodes and edges in the following way:

$$\begin{aligned} &\mathscr{G} = (\mathscr{N}, \mathscr{E}) \\ &\mathscr{N} := \{0, \dots, |\mathscr{N}| - 1\} \\ &\mathscr{E} \subseteq \{(n, m) \mid n, m \in \mathscr{N}, n \neq m\} \end{aligned} \tag{1}$$

where $\mathscr{G}$ is the undirected graph, $\mathscr{N}$ denotes the set of nodes and $\mathscr{E}$ denotes the set of edges. The number of nodes and edges are represented by $|\mathscr{N}|$ and $|\mathscr{E}|$ respectively. A schematic of the graph is presented in Figure 1. The nodes $\{39, \dots, 46\}$ are the the 8 medium-voltage rings that connect all the consumption and production to the high-voltage grid. For each of these regions the aggregated consumption and production data is acquired from EDSN. The rest of the nodes are connecting nodes that do not have consumption or production.

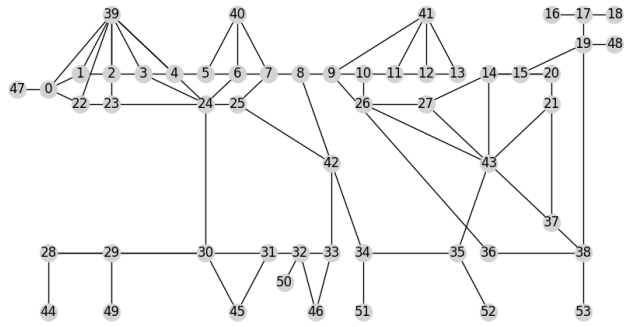


Fig. 1: Graph of the studied high-voltage power grid, where each node represents a coupling substation and each edge represents a transmission line connection between substations.

B. Linearised Dynamical Network Model

The transmission network is represented by a linear time-varying model:

$$x(k+1) = A(k)x(k) + B(k)u(k) + B(k)w(k), \tag{2}$$

where $x(k)$ denotes the current active and reactive power injections and line flows, $u(k)$ the controllable nodal power adjustments, and $w(k)$ the uncontrollable nodal power adjustments defined as

$$\begin{aligned} x(k) = \big[&P^{(1)}(k), \dots, P^{(|\mathscr{N}|)}(k), \\ &Q^{(1)}(k), \dots, Q^{(|\mathscr{N}|)}(k), \\ &P^{(1,2)}(k), \dots, P^{(|\mathscr{E}|)}(k), \\ &Q^{(1,2)}(k), \dots, Q^{(|\mathscr{E}|)}(k)\big]^\mathsf{T}, \\ u(k) = \big[&\Delta P_u^{(1)}(k), \dots, \Delta P_u^{(|\mathscr{N}|)}(k), \\ &\Delta Q_u^{(1)}(k), \dots, \Delta Q_u^{(|\mathscr{N}|)}(k)\big]^\mathsf{T}, \\ w(k) = \big[&\Delta P_w^{(1)}(k), \dots, \Delta P_w^{(|\mathscr{N}|)}(k), \\ &\Delta Q_w^{(1)}(k), \dots, \Delta Q_w^{(|\mathscr{N}|)}(k)\big]^\mathsf{T}, \end{aligned} \tag{3}$$

where $P^{(n)}$ and $Q^{(n)}$ denote the real and reactive power at node n, $P^{(n,m)}$ and $Q^{(n,m)}$ denote the real and reactive power through transmission line (n, m), $\Delta P_u^{(1)}$, $\Delta Q_u^{(1)}$ denote the controlled real and reactive power change at node n, and $\Delta P_w^{(1)}$, $\Delta Q_w^{(1)}$ denote uncontrolled real and reactive power change at node n. The state dynamics of nodal powers are defined as follows:

$$\begin{aligned} P^{(n)}(k+1) = P^{(n)}(k) + \Delta P_u^{(n)}(k) + \Delta P_w^{(n)}(k) \\ \forall n \in \mathscr{N} \\ Q^{(n)}(k+1) = Q^{(n)}(k) + \Delta Q_u^{(n)}(k) + \Delta Q_w^{(n)}(k) \\ \forall n \in \mathscr{N}. \end{aligned} \tag{4}$$

The transmission line flows are linearised using first-order Taylor expansion of the AC power flow equations [8]:

$$\begin{aligned} \Delta P^{(n,m)}(k) = &\sum_{l \in \mathscr{N}} g_{\text{pp}}^{(n,m),l}(k) \Delta P^l(k) \\ &+ g_{\text{pq}}^{(n,m),l}(k) \Delta Q^l(k) \quad \forall (n,m) \in \mathscr{E} \\ \Delta Q^{(n,m)}(k) = &\sum_{l \in \mathscr{N}} g_{\text{qp}}^{(n,m),l}(k) \Delta P^l(k) \\ &+ g_{\text{qq}}^{(n,m),l}(k) \Delta Q^l(k) \quad \forall (n,m) \in \mathscr{E}, \end{aligned} \tag{5}$$

where $g_{\text{pp}}^{(n,m),l}(k)$ are Alternating Current (AC) power transfer distribution factors from real power change in node l to real power transmission in line (n, m) similar to the work in [8]. The model in (2) is for a single time step but can easily be adapted to include all time steps in the prediction horizon in the following way

$$\boldsymbol{x}(k+1) = S(k)x(k) + T(k)\boldsymbol{u}(k) + T(k)\boldsymbol{w}(k) \tag{6}$$

where the matrices $S(k)$ and $T(k)$ are defined as

$$\begin{aligned} S(k) &= \begin{bmatrix} I_{(2|\mathscr{N}|+2|\mathscr{E}|)\times(2|\mathscr{N}|+2|\mathscr{E}|)} \\ \vdots \\ I_{(2|\mathscr{N}|+2|\mathscr{E}|)\times(2|\mathscr{N}|+2|\mathscr{E}|)} \end{bmatrix}, \\ T(k) &= \begin{bmatrix} B(k) & 0 & 0 \\ \vdots & \ddots & 0 \\ B(k) & \dots & B(k) \end{bmatrix} \end{aligned} \tag{7}$$

with

$$\begin{aligned}
\boldsymbol{x}(k+1) &= \begin{bmatrix} x(k+1) \\ x(k+2) \\ \vdots \\ x(k+n_{\mathrm{p}}) \end{bmatrix}, \\
\boldsymbol{u}(k) &= \begin{bmatrix} u(k) \\ u(k+1) \\ \vdots \\ u(k+n_{\mathrm{p}}-1) \end{bmatrix}, \\
\boldsymbol{w}(k) &= \begin{bmatrix} w(k) \\ w(k+1) \\ \vdots \\ w(k+n_{\mathrm{p}}-1) \end{bmatrix}, \\
B(k) &= \begin{bmatrix} I_{|\mathscr{N}|\times|\mathscr{N}|} & 0 \\ 0 & I_{|\mathscr{N}|\times|\mathscr{N}|} \\ G_{\mathrm{pp}}(k) & G_{\mathrm{pq}}(k) \\ G_{\mathrm{qp}}(k) & G_{\mathrm{qq}}(k) \end{bmatrix},
\end{aligned} \tag{8}$$

where the bold symbols indicate that the quantities represent vectors over the entire prediction horizon n_{p} and $G_{\mathrm{pp}}(k), G_{\mathrm{pq}}(k), G_{\mathrm{qp}}(k), G_{\mathrm{qq}}(k)$ are defined as in (36). To ensure the validity of the linearisation, small voltage angle differences are required ($< 3.5\%$), which is appropriate for high-voltage networks [9]. The apparent power transmission limit for a transmission line, commonly referred to as the thermal limit, defines the maximum apparent power transfer. It is formulated as a quadratic constraint on the apparent power magnitude as follows [10], [11]:

$$\begin{aligned}
&S^{(n,m)}(k+1) = \left(P^{(n,m)}(k+i)\right)^2 + \left(Q^{(n,m)}(k+i)\right)^2 \\
&S^{(n,m)}(k+1) \leq \left(S^{\max}\right)^2 \; \forall i \in \{1,\ldots,n_{\mathrm{p}}\}, (n,m) \in \mathscr{E},
\end{aligned} \tag{9}$$

where $S^{\max}$ denotes the maximum apparent power transferable through the transmission lines, and $P^{(n,m)}(k+i)$ and $Q^{(n,m)}(k+i)$ represent the corresponding real and reactive power flows at time step $k+i$.

C. Market Model

The Dutch congestion market operates as a pool-based system where participants submit flexibility offers. Two types of offers, the Profile Offer (PO) and Flex-Time Offer (FTO) are developed based on the requirements formulated in [12].

1) Profile Offers: POs are only valid at a specific moment in time and must specify a fixed power profile for each time step. Such offers are represented by the tuple

$$\mathrm{PO}^{(\mathrm{o})} : (n, t^{\mathrm{start},(\mathrm{o})}, t^{\mathrm{stop},(\mathrm{o})}, \beta^{\min,(\mathrm{o})}, P^{(\mathrm{o})}, c^{(\mathrm{o})}, m^{(\mathrm{o})}), \tag{10}$$

where n denotes the node index associated with offer o. The start and end time of the offer are denoted by $t^{\mathrm{start},(\mathrm{o})}$ and $t^{\mathrm{stop},(\mathrm{o})}$ respectively. The minimum activation fraction is defined as $\beta^{\min,(\mathrm{o})}$. The power profile associated with offer o, denoted by $P^{(\mathrm{o})}$, defines the amount flexible power at each time step within the offer's start and end time. Finally, the price per volume is denoted by $c^{(\mathrm{o})}$ and market direction by $m^{(\mathrm{o})} \in \{-1, 1\}$ with $+1$ for a buy bid and -1 for a sell bid. The power profile $P^{(\mathrm{o})}$ is defined as

$$P^{(\mathrm{o})} = [P^{(n),(\mathrm{o})}(t^{\mathrm{start},(\mathrm{o})}) \ldots P^{(n),(\mathrm{o})}(t^{\mathrm{stop},(\mathrm{o})})] \tag{11}$$

2) Flex-Time Offers: FTOs are offers with a constant maximum power but a flexible activation period, allowing the market mechanism to optimally allocate its operation in time. This offer is defined by the tuple

$$\begin{aligned}
\mathrm{FTO}^{(\mathrm{o})} : (&n, t^{\mathrm{start},(\mathrm{o})}, t^{\mathrm{stop},(\mathrm{o})}, \ell^{\min,(\mathrm{o})}, \ell^{\max,(\mathrm{o})}, \beta^{\min,(\mathrm{o})}, \\
&P^{\max,(\mathrm{o})}, c^{(\mathrm{o})}, m^{(\mathrm{o})}),
\end{aligned} \tag{12}$$

where n denotes the node index associated with offer o. The set of all FTO offers is denoted by $\mathcal{O}^{\mathrm{FTO}}$ and each element $\mathrm{FTO}^{\mathrm{o}} \in \mathcal{O}^{\mathrm{FTO}}$ represents a single FTO. The parameters $t^{\mathrm{start},(\mathrm{o})}$ and $t^{\mathrm{stop},(\mathrm{o})}$ define the earliest activation and latest possible end time, while $\ell^{\min,(\mathrm{o})}$ and $\ell^{\max,(\mathrm{o})}$ specify the minimum and maximum consecutive activation periods. The variables $\beta^{\min,(\mathrm{o})}$ and $P^{\max,(\mathrm{o})}$ denote the minimum activation fraction and maximum power quantity. The offer cost is denoted by $c^{\mathrm{offer},(\mathrm{o})}$, and $m^{\mathrm{offer},(\mathrm{o})}$ indicates the market direction, with $+1$ for buy and -1 for sell offers.

III. Data Usage and Uncertainty

This section presents the derivation of an uncertainty model as well as the out-of-sample forecasting results used to represent uncertainty in the proposed CC-MPC–RL framework.

A. Data Description

Eight Dutch regions are considered, each featuring both electricity production and consumption. The data, provided by EDSN, include all distribution-level connections but exclude those above 60 MW, such as conventional power plants directly connected to the transmission grid. The missing data is filled by assumping power balance and allocating the missing generation capacityx to representative generator nodes.

Consumption exhibits highly regular daily and weekly patterns. As these are well predicted in existing operational forecasts, real measured consumption values are used directly. The statistical modelling effort focuses on renewable generation, which is much more volatile and constitutes the main source of uncertainty.

B. Forecasting Model and Results

Renewable generation in each region is modelled using a SARIMA process. For Noord-Holland, the best-fitting model is $\mathrm{SARIMA}(5,0,0,0,1,1,96)$, capturing both short-term correlations and daily seasonality. Similar structures were obtained for other regions.

Residual diagnostics confirm good model quality:

- Residuals are centred around zero with no systematic bias;
- Their distribution is approximately Gaussian with moderately heavy tails;
- Residual autocorrelation lies within statistical bounds, showing that temporal dependencies are well captured.

Some heteroscedasticity, i.e., time-varying variance, remains due to weather-driven variations in renewable output, motivating the adaptive variance treatment later used in the CC-MPC–RL framework.

Out-of-sample prediction performance is evaluated for horizons consistent with the control design ($n_\mathrm{p} = 16\}$). Table I summarises the results for $n_\mathrm{p} = 16$ in terms of Root Mean Square Error (RMSE) and Mean Absolute Precentage Error (MAPE), compared to the realised generation. Most regions achieve MAPE values between 10–30%, which is acceptable given the inherent variability of aggregated RESs. Outliers such as Groningen and Friesland show high MAPE due to periods of very low production, which inflate percentage errors despite moderate absolute deviations. Overall, the results show that

Area	RMSE (mean)	MAPE	MAPE (variance)
Noord-Holland	6497 [kW]	13.86%	8.64%
Zuid-Holland	26608 [kW]	23.25%	20.02%
Groningen	22882 [kW]	495.45%	4959.55%
Zeeland	11765 [kW]	31.69%	32.00%
Brabant	17001 [kW]	19.89%	18.80%
Limburg	9965 [kW]	23.03%	19.24%
Friesland	7679 [kW]	41.94%	53.05%
Utrecht, Flevopolder, Gelderland	19864 [kW]	27.68%	27.82%

TABLE I: Out-of-sample prediction accuracy per region for $n_\mathrm{p} = 16$.

the fitted models capture key statistical properties of renewable generation while maintaining acceptable forecast accuracy for use in stochastic control.

IV. Chance-constrained MPC and RL-based adaptation

This section introduces the proposed market-based, sample-approximated CC-MPC formulation for CM.

A. Flex-Time Offers

The behaviour of the FTOs is described by the following constraints, which employ time-dependent binary activation variables $\delta^{(\mathrm{o})}(k)$:

$$\begin{aligned}
\Delta P_u^{(n),(\mathrm{o})}(k+i) \leq & M^{\mathrm{high}}\delta^{(\mathrm{o})}(k+i) \\
& - \sum_{j=t^{\mathrm{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(j) \\
& \forall i \in \{1,\dots,n_\mathrm{p}\} \\
\Delta P_u^{(n),(\mathrm{o})}(k+i) \geq & M^{\mathrm{low}}\delta^{(\mathrm{o})}(k+i) \\
& - \sum_{j=t^{\mathrm{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(j) \\
& \forall i \in \{1,\dots,n_\mathrm{p}\} \\
\Delta P_u^{(n),(\mathrm{o})}(k+i) \leq & \beta^{(\mathrm{o})} P^{\mathrm{max,(o)}} \\
& - \sum_{j=t^{\mathrm{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(j) \\
& - M^{\mathrm{low}}(1-\delta^{(\mathrm{o})}(k+i)) \\
& \forall i \in \{1,\dots,n_\mathrm{p}\} \\
\Delta P_u^{(n),(\mathrm{o})}(k+i) \geq & \beta^{(\mathrm{o})} P^{\mathrm{max,(o)}} \\
& - \sum_{j=t^{\mathrm{start}}}^{k+i-1} \Delta P_u^{(n),(\mathrm{o})}(j) \\
& - M^{\mathrm{high}}(1-\delta^{(\mathrm{o})}(k+i)) \\
& \forall i \in \{1,\dots,n_\mathrm{p}\},
\end{aligned} \tag{13}$$

where M^{high} and M^{low} are large positive and negative constants, respectively, bounding the feasible range of $\Delta P_u^{(n)}(k+i)$. This formulation ensures that the power adjustment $\Delta P_u^{(n)}(k+i)$ assumes meaningful values only when the offer is active (i.e., $\delta^{(\mathrm{o})}(k+i) = 1$), while it is forced to zero when the offer is inactive ($\delta^{(\mathrm{o})}(k+i) = 0$). The minimum activation duration is enforced as follows:

$$\begin{aligned}
\sum_{j=i}^{k+\ell^{\min}-1} \delta^{(\mathrm{o})}(k+j) \geq \ell^{\min}\big(\delta^{(\mathrm{o})}(k+i) - \delta^{(\mathrm{o})}(k+i-1)\big) \\
\forall k+i \in \{t^{\mathrm{start}},\dots,t^{\mathrm{stop}}-\ell^{\min}+1\},
\end{aligned} \tag{14}$$

which guarantees that when the activation variable switches from 0 at time $k+i-1$ to 1 at time $k+i$, the offer remains active for at least $\ell^{\min}$ consecutive time steps. In other words, each activation must last no shorter than the minimum duration specified in the FTO.

To limit the maximum activation duration, an upper bound $\ell^{\max}$ is imposed:

$$\sum_{j=t^{\mathrm{start}}}^{t^{\mathrm{stop}}} \delta_j^{(\mathrm{o})} \leq \ell^{\max}, \tag{15}$$

which restricts the total number of time periods during which the offer can remain active to the specified limit. However, this formulation still allows multiple non-consecutive activations

within the activation window. To prevent such behaviour, the following constraint is introduced:

$$\sum_{j=i}^{t^{\text{stop}}} \delta_j^{(\text{o})} \leq \ell^{\max}\big(1+\delta^{(\text{o})}(k+i)-\delta^{(\text{o})}(k+i-1)\big) \quad \forall k+i \in \{t^{\text{start}}+\ell^{\min},\dots,t^{\text{start}}+\ell^{\max}-1\}, \tag{16}$$

which ensures that once the activation variable transitions from 1 at time $k+i-1$ to 0 at time $k+i$, all subsequent activations are prohibited within the same time window.

B. Profile Offers

In the Dutch congestion market, only real-power offers are currently traded; therefore, only these are considered here, although the proposed model can also accommodate reactive-power offers. The behaviour of the POs is modelled through the following set of constraints, employing the binary activation variable $\delta^{(\text{o})}$:

$$\begin{aligned}
\Delta P_u^{(n),(\text{o})}(k+i) &\leq M^{\text{high}}\delta^{(\text{o})} - \sum_{j=t^{\text{start}}}^{k+i-1} \Delta P_u^{(n),(\text{o})}(j) \quad \forall i \in \{1,\dots,n_{\text{p}}\}\\
\Delta P_u^{(n),(\text{o})}(k+i) &\geq M^{\text{low}}\delta^{(\text{o})} - \sum_{j=t^{\text{start}}}^{k+i-1} \Delta P_u^{(n),(\text{o})}(k+j)\\
\Delta P_u^{(n),(\text{o})}(k+i) &\leq \beta^{(\text{o})} P^{(n),(\text{o})}(k+i) \quad \forall i \in \{1,\dots,n_{\text{p}}\}\\
&\quad - \sum_{j=t^{\text{start}}}^{k+i-1} \Delta P_u^{(n),(\text{o})}(k+j) - M^{\text{low}}(1-\delta^{(\text{o})}) \quad \forall i \in \{1,\dots,n_{\text{p}}\}\\
\Delta P_u^{(n),(\text{o})}(k+i) &\geq \beta^{(\text{o})} P^{(n),(\text{o})}(k+i) - \sum_{j=t^{\text{start}}}^{k+i-1} \Delta P_u^{(n),(\text{o})}(k+j)\\
&\quad - M^{\text{high}}(1-\delta^{(\text{o})}) \quad \forall i \in \{1,\dots,n_{\text{p}}\},
\end{aligned} \tag{17}$$

where M^{high} and M^{low} denote sufficiently large positive and negative constants defining the upper and lower bounds of the feasible range for $\Delta P_u^{(n),(\text{o})}(k)$. When the offer is inactive ($\delta^{(\text{o})} = 0$), the inequalities enforce $\Delta P_u^{(n),(\text{o})}(k+i) = -\sum_{j=t^{\text{start}}}^{k+i-1} \Delta P_u^{(n),(\text{o})}(j)$ to ensure that it is zero before activation, and the first step after activation the output is forced back its original value. Then the sum of all actions is forced to zero. Conversely, when $\delta^{(\text{o})} = 1$, the equations permit $\Delta P_u^{(n),(\text{o})}(k)$ to take values consistent with the desired power change determined by $P_u^{(n),(\text{o})}(k)$ and $\beta^{(\text{o})}$.

C. Balancing requirement

The overall control input can be decomposed into the contributions of individual offers. Accordingly, the total control action at time step k is expressed as

$$u(k+i) = \sum_{\text{o}\in\mathcal{O}} u^{(\text{o})}(k+i) \quad \forall i \in \{0,\dots,n_{\text{p}}-1\}, \tag{18}$$

where $u^{(\text{o})}(k+i)$ denotes the control action associated with the offer indexed by o. This quantity is zero for all nodes except the one where the offer is placed, as defined by

$$u^{(\text{o})}(k+i) = \begin{cases} \big[0\dots 0\ \Delta P_u^{(n),(\text{o})}(k+i)\ 0\dots 0\big]^T \\ \qquad \forall k+i \in \{t^{\text{start,(o)}},\dots,t^{\text{stop,(o)}}\} \\ \big[0\dots 0\big]^T \quad \text{otherwise,} \end{cases} \tag{19}$$

where $\Delta P_u^{(n)}(k+i)$ represents the power adjustments associated with offer (o) as detailed in Subsections IV-B and IV-A.

The net effect of CM actions must not alter the overall system power balance. This requirement is formulated as

$$\sum_{\text{o}\in\mathcal{O}} u^{(\text{o})}(k+i) = 0 \quad \forall i \in \{0,\dots,n_{\text{p}}-1\}, \tag{20}$$

ensuring that the aggregate control actions across all offers remain power-neutral at all times.

Furthermore, once an offer is activated, it must remain active throughout the prediction horizon. Consequently, the total control action applied at time step k is the cumulative sum of all past n_{p} control actions, expressed as

$$\boldsymbol{u}(k) = \sum_{i=0}^{n_{\text{p}}} \boldsymbol{u}(k|k-i) \tag{21}$$

where $\boldsymbol{u}(k|k-i)$ denotes the sequence of control actions computed at time step $k-i$, temporally shifted to align with the current time step and zero-padded as necessary.

D. Dynamical model

The dynamical model in (6) is adapter for the sample approximated CC-MPC in the following way

$$\boldsymbol{x}^{(\text{s})}(k+1) = S(k)x(k) + T(k)\boldsymbol{u}(k) + T(k)\boldsymbol{w}^{(\text{s})}(k) \quad \forall \text{s} \in \{1,\dots,n_{\text{s}}\} \tag{22}$$

where $\boldsymbol{x}^{(s)}(k)$ is the state trajectory for each scenario s, $\boldsymbol{u}(k)$ the sum of the past control actions as defined in (21), and $\boldsymbol{w}^{(s)}(k)$ a disturbance trajectory generated by the SARIMA model from III.

E. Limits

Thermal limit on apparent power flow, as defined in (9), is the only quadratic constraint. To reduce the problem from a mixed-integer second-order cone program into a mixed-integer linear program, a linear approximation is adopted by assuming a high power factor (> 0.95), such that $P^{(n,m)}(k+i) \approx S^{(n,m)}(k+i)$ [13]. Under this assumption, the apparent power limit can be expressed as

$$|P^{(n,m)}(k+i)| \leq S^{\max}. \tag{23}$$

This deterministic constraint is then replaced by a scenario-based approximation of the stochastic constraint:

$$\frac{1}{n_{\mathrm{s}}}\sum_{s=1}^{n_{\mathrm{s}}}\mathbb{1}\left(|P^{(n,m),(\mathrm{s})}(k+i)|-S^{\max}\geq 0\right)\leq 1-\alpha, \quad (24)$$

where $\mathbb{1}(\cdot)$ equals 1 if the constraint is violated and 0 otherwise, and $\alpha \in [0,1)$ defines the required risk level. Since constraint satisfaction cannot be guaranteed (even with chance-constraints), binary variables $\delta^{(\mathrm{s})}$ and shared slack variables $z^{(n,m)}(k+i)$ are introduced:

$$|P^{(n,m),(\mathrm{s})}(k+i)|-S^{\max}\leq z^{(n,m)}(k+i)+M\delta^{(\mathrm{s})}, \quad (25)$$

with $z^{(n,m)}(k+i)$ capturing the maximum violation across all scenarios. These slack variables are penalised in the objective as follows

$$\begin{aligned} &c_{\mathrm{s}}\max\Big(\sum_{s=1}^{n_{\mathrm{s}}}\delta^{(\mathrm{s})}-(1-\alpha)n_{\mathrm{s}},0\Big)\\ &+\sum_{(n,m)\in\mathscr{E}}\sum_{i=1}^{n_{\mathrm{p}}}c_z z^{(n,m)}(k+i), \end{aligned} \quad (26)$$

where c_{s} ensures that additional scenarios with violations incur a higher cost than the corresponding slack magnitude. This formulation penalises excessive violations while tolerating those permitted by the risk parameter. Additionally, nodal power constraints are imposed to ensure that both real and reactive power injections remain within operational limits:

$$\begin{aligned} |P^{(n)}(k+i)|&<P^{(n),\max} \quad \forall i\in\{1,\dots,n_{\mathrm{p}}\}, n\in\mathscr{N}\\ |Q^{(n)}(k+i)|&<Q^{(n),\max} \quad \forall i\in\{1,\dots,n_{\mathrm{p}}\}, n\in\mathscr{N}. \end{aligned} \quad (27)$$

F. Cost

The total cost for the network operator is given by the spread between the buy and sell offers, defined as

$$C^{\text{total}}=\sum_{\mathrm{o}\in\mathcal{O}}m^{(\mathrm{o})}\beta^{(\mathrm{o})}c^{(\mathrm{o})}\sum_{i=1}^{n_{\mathrm{p}}}\sum_{j=0}^{i}u^{(\mathrm{o})}_{k+j}, \quad (28)$$

where $c^{(\mathrm{o})}$ is the offer price and $m^{(\mathrm{o})}$ indicates the buy/sell direction.

G. Full dynamical model

Combining all components, the complete optimisation problem is formulated as

$$\begin{aligned} \min_{\boldsymbol{x}^{\mathrm{s}}(k+1),\boldsymbol{u}(k|k),\boldsymbol{\delta},\boldsymbol{\beta}}\quad & (26)+(28)\\ \text{s.t.}\quad & (19)-(22)\\ & (25)\\ & (27)\\ \text{PO}^{(\mathrm{o})}:\quad & (17)\\ \text{FTO}^{(\mathrm{o})}:\quad & (13)-(16), \end{aligned} \quad (29)$$

where $\boldsymbol{\delta}$ is the collection of $\delta^{(\mathrm{o})}$ variables and $\boldsymbol{\beta}$ is the collection of all $\beta^{(\mathrm{o})}$ variables.

V. Reinforcement Learning-Enhanced CC–MPC

The proposed method integrates RL with the CC-MPC formulation in (29) to improve robustness and adaptability under time-varying uncertainties. An RL-agent dynamically scales the variance of the innovations in the SARIMA-based disturbance models, inspired by the adaptive robustification concept introduced by [14]. By learning this uncertainty scaling online, the controller balances feasibility and performance in response to real-time disturbances.

Figure 2 illustrates the overall control architecture. The true system provides the state feedback $x(k)$ and disturbance $w(k)$ to the AutoRegressive Moving-Average (ARMA) models, which predicts the next n_{p} values of $w^{\mathrm{s}}(k+i)$ for n_{s} scenarios, with $i\in\{1,\dots,n_{\mathrm{p}}\}$. The RL-agent observes the system behaviour and recent performance to output a set of scaling factors $\kappa_b \forall b\in\{1,\dots,8\}$, one per region, that adjust the innovation variance. These predictions are passed to the CC-MPC, which computes the optimal control input $u(k)$, closing the loop and continuously updating the RL policy for adaptive uncertainty handling.

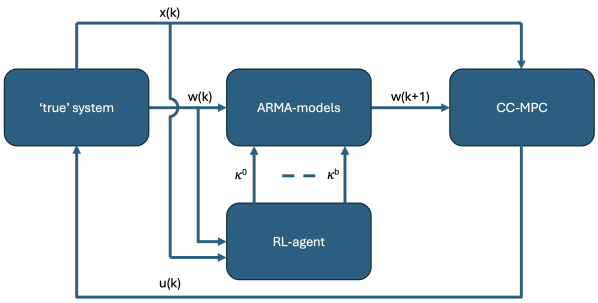


Fig. 2: Overview of the proposed closed-loop framework integrating RL with CC-MPC. The RL-agent adjusts the innovation variance in the ARMA-based disturbance predictions to enhance robustness and adaptability.

A. RL Formulation

The agent's state vector $s(k)$ comprises the system state $x(k)$ and recent disturbance history:

$$s(k)=[x(k)^T w(k)^T w(k-1)^T\dots w(k-191)^T]^T. \quad (30)$$

The action $a(k)$ defines the scaling factors κ_b applied to the innovation variance of each SARIMA model,

$$\begin{aligned} &\varepsilon_b(k+1)\sim\mathcal{N}(0,\,\kappa_b\sigma^2),\forall b\in\{1,\dots,8\}\\ &\text{with } \kappa_b\in\{0.5,\dots,2.0\} \end{aligned} \quad (31)$$

The immediate reward penalises both operational cost and constraint violations:

$$r(k)=-c^{\text{total}}C^{\text{total}}(a(k),s(k))-c^{\text{vio}}C^{\text{vio}}(s(k+1)), \quad (32)$$

where C^{total} is defined as in (28) and is the sum of all constraint violations squared:

$$C_{\text{vio}}(s(k+1)) = \sum_{(n,m)\in\mathscr{E}} \max(0, |P^{(n,m)}(k+1)| - S^{\max})^2. \tag{33}$$

B. Learning Framework

The policy $\pi(a|s)$ is trained using a Double Dueling Deep Q-Network (DQN) with an action-branching architecture [15]. Each branch b independently estimates the advantage of its local action, while a shared state-value stream captures common information:

$$\begin{aligned} Q^{(b)}(s(k), a^{(b)}(k)) =& V(s(k)) + A^{(b)}(s(k), a^{(b)}(k)) \\ &- \frac{1}{|\mathcal{A}^{(b)}|} \sum_{a^{(b)}\in\mathcal{A}^{(b)}} A^{(b)}(s(k), a^{(b)}), \end{aligned} \tag{34}$$

for each branch $b \in \{1, \ldots, 8\}$. The temporal-difference loss for training is

$$L(\theta) = \mathbb{E}\left[\frac{1}{8}\sum_{b=1}^{8}\left(y_t^{(b)} - Q^{(b)}(s(k), a^{(b)}(k); \theta)\right)^2\right], \tag{35}$$

with Double Q-learning targets to reduce overestimation bias. Experience replay is used to decorrelate samples and improve stability. The overall training procedure is summarised in Algorithm 1.

Init: Initialise Q_θ, target $Q_{\bar{\theta}} \leftarrow Q_\theta$, replay buffer $\mathcal{D}$
while *training* **do**
  Observe state $s(k)$ and select $a(k)$ via ε-greedy policy;
  Apply κ_b to SARIMA models and generate $\boldsymbol{w}^{\text{s}}(k)$ and compute $u(k)$ using (29);
  Execute $u(k)$ on the plant to obtain $s(k+1)$ and reward $r(k)$;
  Store $(s(k), a(k), r(k), s(k+1))$ in $\mathcal{D}$ and update θ using $L(\theta)$;
  Every τ steps: update target $Q_{\bar{\theta}} \leftarrow Q_\theta$;
  $k = k + 1$;
end

Algorithm 1: Training of the RL–enhanced CC–MPC controller

The cumulative reward evolution during training is shown in Figure 3. It exhibits oscillatory behaviour and does not converge. Several factors may explain this outcome. First, constraint satisfaction is not always possible due to limitations on the available actions, which results in continuous penalties. Second, the model may not have been trained for a sufficient number of episodes. Finally, the presence of multiple penalties may have introduced conflicting objectives that hinder convergence.

VI. Case Study and Results

This section evaluates the proposed CM strategy introduced in Section V on the Dutch high-voltage grid. First, the simulation set-up, data usage, and offer generation procedure is described. Then, a single simulation is highlighted, followed by a quantitative comparison of all methods across multiple operating scenarios. The analysis focuses on four design dimensions: prediction horizon, forecast quality, safety parameter α in the CC-MPC, and market flexibility (number of offers). Finally,the impact of the RL-based enhancement of the CC-MPC is discussed.

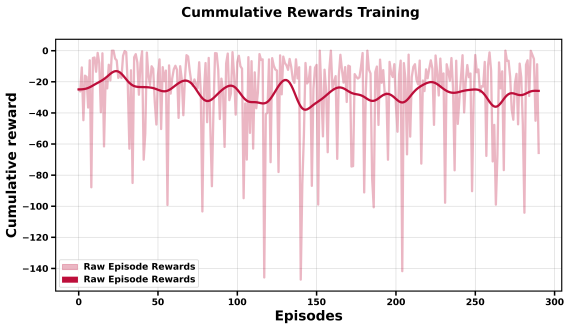


Fig. 3: Cumulative rewards during reinforcement learning training.

A. Simulation and Experimental Setup

The proposed control strategy is tested under varying modelling and operational conditions and compared to different methods. The following control schemes are compared: (i) Algorithmic Greedy Matching, (ii) nominal Model Predictive Control (MPC), (iii) chance-constrained MPC (CC-MPC), and (iv) RL-enhanced CC-MPC (CC-MPC-RL). All controllers are evaluated in terms of constraint satisfaction (total violation magnitude and violation count) and operational cost, and benchmarked against two reference cases: (i) an uncontrolled scenario without CM actions and (ii) the idealised MPC-PP using perfect future knowledge.

Two different prediction horizon lengths are considered. A longer horizon allows the inclusion of longer admissible offers, thereby increasing the feasibility of offer matching and overall system flexibility. However, for linearised network models, it also amplifies the accumulation of model mismatch away from the operating point and increases optimisation complexity.

The safety parameter $\alpha \in [0, 1)$ in the chance constraints is used to control the trade-off between robustness and performance in the CC-MPC. Two values of α are tested to assess how more conservative (higher α) or more aggressive (lower α) operation affects violations and cost.

To quantify the effect of forecast quality, the MPC is evaluated under two prediction settings: (i) stochastic forecasts generated by the SARIMA models introduced in Section III, and (ii) perfect disturbance information (MPC-PP). This comparison isolates the impact of forecast errors on closed-loop performance.

Finally, market flexibility is varied by changing the number of available offers. Four offer sets are analysed, combining two prediction horizon lengths with two offer quantities (100

and 200 offers per horizon). This allows us to study how the availability of flexibility impacts congestion mitigation and costs.

B. Data Usage

Table II summarises the partitioning of the historical dataset across modelling, training, and evaluation. The period from 2023-04-20 to 2024-04-20 is used to fit the ARMA models developed in Section III, which provide the stochastic forecasts required by the predictive controllers.

The subsequent year, 2024-04-20 to 2025-04-20, is used both for training the RL-based enhancement and for evaluating all control strategies. Specifically, 85% of this data is allocated to RL training to expose the agent to a broad range of operating conditions, while the remaining 15% is reserved for testing and producing the final simulation results. The 15% test subset is randomly sampled and corresponds to approximately 50 simulation days. This split ensures that controller design (forecast modelling and RL policy learning) and performance evaluation are performed on disjoint data, providing an unbiased assessment of the proposed methods.

Data range	Amount	Usage
2023-04-20 to 2024-04-20	100%	Training ARMA models
2024-04-20 to 2025-04-20	85%	Training RL
2024-04-20 to 2025-04-20	15%	Producing results

TABLE II: Data usage for forecasting, RL training, and performance evaluation.

for $i = \{1, \dots, N_{offers}\}$ **do**
  $n \sim \{39, 40, 41, 42, 43, 44, 45, 46\}$;
  $m^{(\mathrm{o})} \sim \{1, -1\}$;
  if $m^{(o)} = 1$ **then**
    $c^{(\mathrm{o})} \sim \mathcal{N}(180, 40)$
  else
    $c^{(\mathrm{o})} \sim \mathcal{N}(220, 50)$
  end
  $t^{\mathrm{start},(\mathrm{o})} \leftarrow \mathrm{UNIF}(0, 96 - n_{\mathrm{p}})$;
  $\ell^{\mathrm{min},(\mathrm{o})} \leftarrow \mathrm{UNIF}(0, n_{\mathrm{p}})$;
  $\ell^{\mathrm{max},(\mathrm{o})} \leftarrow \mathrm{UNIF}(\ell^{\mathrm{min},(\mathrm{o})}, n_{\mathrm{p}})$;
  $t^{\mathrm{stop},(\mathrm{o})} \leftarrow t^{\mathrm{start},(\mathrm{o})} + \ell^{\mathrm{max},(\mathrm{o})}$;
  $\beta^{\mathrm{min},(\mathrm{o})} \leftarrow \mathrm{UNIF}(0, 1)$;
  if *FTO*$^{(o)}$ **then**
    $P^{\mathrm{max},(\mathrm{o})} \leftarrow \mathcal{N}(m^{(\mathrm{o})}30, 10)$;
    $\mathcal{O} \leftarrow (n, t^{\mathrm{start},(\mathrm{o})}, t^{\mathrm{stop},(\mathrm{o})}, \ell^{\mathrm{min},(\mathrm{o})}, \ell^{\mathrm{max},(\mathrm{o})}, \beta^{\mathrm{min},(\mathrm{o})}, P^{\mathrm{max},(\mathrm{o})}, c^{(\mathrm{o})}, m^{(\mathrm{o})})$;
  else
    $P^{1,(\mathrm{o})} \leftarrow \mathcal{N}(m^{(\mathrm{o})}30, 10)$;
    $P^{2,(\mathrm{o})} \leftarrow \mathcal{N}(m^{(\mathrm{o})}30, 10)$;
    $P^{\mathrm{max},(\mathrm{o})} \leftarrow [P^{1,(\mathrm{o})} \dots P^{1,(\mathrm{o})}\ P^{2,(\mathrm{o})} \dots P^{2,(\mathrm{o})}]$;
    $\mathcal{O} \leftarrow (n, t^{\mathrm{start},(\mathrm{o})}, t^{\mathrm{stop},(\mathrm{o})}, \beta^{\mathrm{min},(\mathrm{o})}, P^{(\mathrm{o})}, c^{(\mathrm{o})}, m^{(\mathrm{o})})$;
  end
end

Algorithm 2: Offer Set Generation

C. Offer Generation and Analysis

To emulate a realistic and diverse flexibility market, the offer set generation algorithm randomly assigns attributes to each offer as defined in (12) and (II-C1). The procedure is summarised in Algorithm 2. It samples bus locations, directions (up- or down-regulation), prices, timing, duration, and power limits from appropriate distributions to create a heterogeneous pool of offers.

Figure 4 shows the average aggregated flexible power for the four offer sets. The green and red shaded regions indicate the available up- and down-regulation power, respectively, and the grey area denotes the matchable flexibility volume, i.e., the maximum power that can be activated since there are sufficient opposing offers.

Figures 4a and 4b compare two offer set sizes for a prediction horizon of $n_{\mathrm{p}} = 8$. As expected, increasing the number of offers broadens both up- and down-regulation ranges and enlarges the matchable area. Figures 4c and 4d show the same comparison for a longer horizon of $n_{\mathrm{p}} = 16$. In all cases, the larger offer sets provide higher flexibility across the entire horizon.

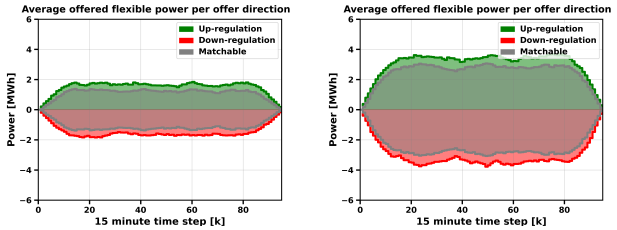
D. Case Study: 2024-08-20

A representative operating day (2024-08-20) is analysed in detail to illustrate the qualitative behaviour of the different control schemes. All methods are tested using the offer set with $n_{\mathrm{p}} = 16$ and 200 flexibility offers. The results are summarised in three figures, highlighting constraint violations, control activations, and aggregated performance metrics.

Figure 5 compares the temporal distribution of constraint violations across all transmission lines. The horizontal axis denotes time in 15-minute intervals; the vertical axis lists the transmission lines. Colour intensity encodes the magnitude of violation, with darker red indicating higher congestion.

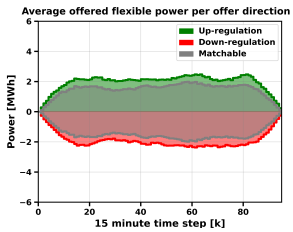
The uncontrolled case exhibits persistent congestion on one structurally overloaded line and additional overloads in the afternoon. The Algorithmic Greedy Matching approach yields similar patterns, as it does not effectively deploy flexibility in this scenario. In contrast, the MPC substantially reduces both the number and severity of violations by actively activating flexibility. The MPC-PP further improves congestion mitigation, confirming that forecast errors contribute to the remaining violations observed under nominal MPC. The chance-constrained controllers, CC-MPC and CC-MPC-RL, provide the most pronounced reduction in congestion, exhibiting more preventive behaviour and mitigating minor overloads. This demonstrates the benefits of probabilistic constraint handling for robustness and reliability.

Figure 6 shows the corresponding control actions. Each subplot displays the total available flexibility (grey area) together with the activated up- (green) and down-regulation (red) power. The Algorithmic method activates no flexibility, consistent with the congestion patterns observed in the heatmap. The MPC-based controllers trigger flexibility in response to pre-

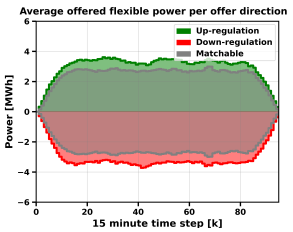


(a) $n_p = 8$, 100 offers

(b) $n_p = 8$, 200 offers



(c) $n_p = 16$, 100 offers



(d) $n_p = 16$, 200 offers

Fig. 4: Aggregated flexible power for each offer set, averaged over all simulation days. The green and red areas denote up- and down-regulation potential; grey denotes matchable flexibility.

dicted or current overloads, with MPC-PP exhibiting somewhat smoother activations due to its perfect knowledge of future disturbances. The CC-MPC and CC-MPC-RL controllers activate flexibility more broadly and more frequently, including in periods where no congestion would have occurred under perfect foresight. This reflects a conservative, preventive strategy that prioritises safety at the expense of higher cost.

Finally, Figure 7 presents the cumulative constraint violations (top panel) and cumulative operational costs (bottom panel). The uncontrolled and Algorithmic cases incur the highest violations, highlighting the inadequacy of simple matching or absence of control. All MPC-based methods significantly reduce violations. The chance-constrained controllers (CC-MPC and CC-MPC-RL) achieve the lowest cumulative violations overall in this specific day, despite relying on imperfect forecasts. However, at significantly increased cost with respect to the MPC-based methods

In terms of cost, the Algorithmic strategy is cheapest because it never activates flexibility, but this is not acceptable from a safety perspective. The MPC and MPC-PP incur higher costs due to the activation of flexibility resources, with the probabilistic methods being most expensive. The increased cost, however, is accompanied by a slight improved congestion prevention under uncertainty, underscoring the trade-off between economic efficiency and robustness.

E. Impact of Prediction Quality

As discussed in Section III, the statistical forecasting models exhibit non-negligible prediction errors. To quantify their impact, the nominal MPC is compared against MPC-PP (perfect predictions) across all combinations of prediction horizon

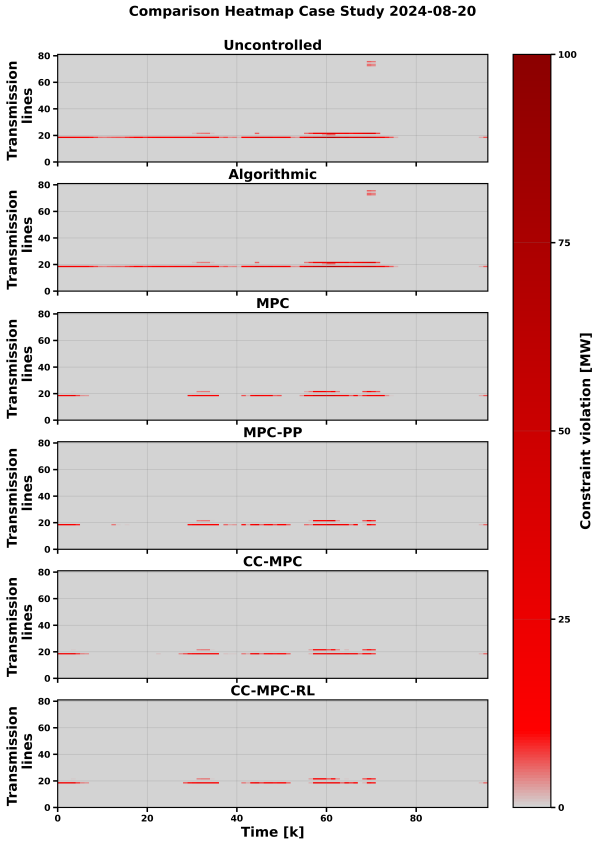


Fig. 5: Heatmap of constraint violations for all control strategies on 2024-08-20. Darker colours indicate higher congestion levels.

and offer set size. Tables III and IV summarise the results, reported relative to the uncontrolled case and averaged over 50 simulation days. Comparing Tables III and IV shows that

MPC	$n_p = 8$ $n_{of} = 100$	$n_p = 8$ $n_{of} = 200$	$n_p = 16$ $n_{of} = 100$	$n_p = 16$ $n_{of} = 200$
Total violation (mean) [%]	-3.91 %	-6.57 %	-35.81 %	-56.18 %
Total violation (variance) [%]	5.97 %	7.74 %	25.0 %	26.78 %
Violation count (mean) [%]	-2.6 %	-4.68 %	-29.22 %	-46.65 %
Violation count (variance) [%]	7.42 %	9.58 %	22.91 %	27.55 %
Cost (mean) [€]	16692.68 €	22527.62 €	302572.72 €	538241.63 €
Cost (variance) [€]	11060.39 €	12117.94 €	219901.91 €	241400.13 €

TABLE III: MPC performance metrics across 50 simulations for varying prediction horizons (n_p) and offer set sizes (n_{of}), relative to the uncontrolled case.

perfect predictions consistently reduce both total violations and violation counts, especially for larger offer sets where the controller can better exploit accurate information. Nonetheless, the relative differences remain moderate, indicating that the nominal MPC retains good congestion mitigation capability even with imperfect SARIMA-based forecasts. Forecast errors

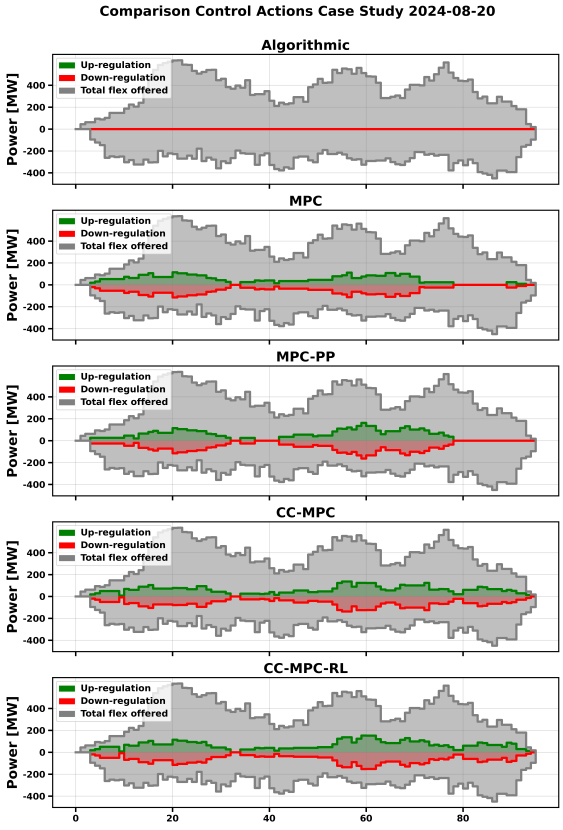


Fig. 6: Control actions for all strategies on 2024-08-20. Grey: available flexibility; green/red: activated up-/down-regulation.

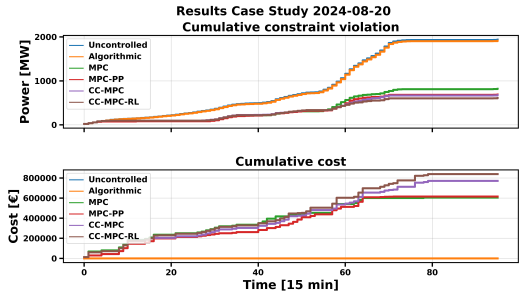


Fig. 7: Cumulative constraint violation (top) and cumulative cost (bottom) for all control strategies on 2024-08-20.

therefore reduce optimality but do not fundamentally compromise closed-loop performance.

F. Impact of the safety Parameter

A key advantage of the CC-MPC framework is the explicit tuning of robustness via the safety parameter α. Table V reports performance metrics for two values of α across all horizon and offer set combinations.

MPC-PP	$n_\mathrm{p} = 8$ $n_\mathrm{of} = 100$	$n_\mathrm{p} = 8$ $n_\mathrm{of} = 200$	$n_\mathrm{p} = 16$ $n_\mathrm{of} = 100$	$n_\mathrm{p} = 16$ $n_\mathrm{of} = 200$
Total violation (mean) [%]	-4.54 %	-7.43 %	-41.49 %	-60.32 %
Total violation (variance) [%]	7.78 %	11.49 %	27.46 %	22.55 %
Violation count (mean) [%]	-1.94 %	-5.0 %	-32.19 %	-48.8 %
Violation count (variance) [%]	6.25 %	10.77 %	25.65 %	21.13 %
Cost (mean) [€]	14042.17 €	22273.09 €	313929.35 €	562637.84 €
Cost (variance) [€]	9827.81 €	14293.88 €	230628.23 €	258304.89 €

TABLE IV: MPC-PP performance metrics across 50 simulations for varying prediction horizons (n_p) and offer set sizes (n_of), relative to the uncontrolled case.

CC-MPC	α	$n_\mathrm{p} = 8$ $n_\mathrm{of} = 100$	$n_\mathrm{p} = 8$ $n_\mathrm{of} = 200$	$n_\mathrm{p} = 16$ $n_\mathrm{of} = 100$	$n_\mathrm{p} = 16$ $n_\mathrm{of} = 200$
Total violation	0.75	-3.03 %	-6.71 %	-39.38 %	-59.2 %
(mean) [%]	0.9	-3.93 %	-7.13 %	-43.77 %	-59.44 %
Total violation	0.75	7.93 %	11.24 %	25.52 %	24.82 %
(variance) [%]	0.9	7.56 %	11.23 %	27.51 %	24.14 %
Violation count	0.75	-1.47 %	-5.06 %	-32.27 %	-50.77 %
(mean) [%]	0.9	-1.88 %	-5.74 %	-35.92 %	-50.85 %
Violation count	0.75	6.38 %	12.87 %	22.9 %	23.43 %
(variance) [%]	0.9	6.02 %	12.64 %	25.61 %	22.51 %
Cost (mean) [€]	0.75	12567.14 €	25642.35 €	355028.02 €	653568.64 €
	0.9	12258.20 €	25623.13 €	371277.57 €	690521.43 €
Cost (variance)	0.75	11445.64 €	16194.16 €	237708.41 €	255827.52 €
[€]	0.9	10249.82 €	17058.65 €	230599.56 €	238974.97 €

TABLE V: CC-MPC performance metrics across 50 simulations for varying n_p, n_of, and safety level α, relative to the uncontrolled case.

Lowering the safety level from $\alpha = 0.9$ to $\alpha = 0.75$ slightly reduces cost but generally increases violations, as expected when relaxing the chance constraints. However, the performance differences are modest. This is largely due to the structural nature of the congestion: many violations are persistent over the operating horizon and cannot be eliminated simply by allowing a lower probability of constraint satisfaction. In such cases, the controller has limited opportunity to exploit the additional safety tolerance, and the gain in economic efficiency remains small.

G. Impact of RL Enhancement

The CC-MPC-RL controller augments the CC-MPC by tuning the variance of the disturbance model based on observed closed-loop performance. In this study, the RL training process did not fully converge: the learning curves displayed oscillatory behaviour and no clear stabilisation to a unique policy (see Section V-A). Nonetheless, the resulting policies can still be evaluated.

Table VI reports performance metrics for CC-MPC-RL. For the short prediction horizon $n_\mathrm{p} = 8$, on which the RL agent was trained, the CC-MPC-RL achieves slightly lower mean violations compared to CC-MPC, indicating that the learned adaptation of uncertainty can provide tangible improvements despite imperfect training. For the longer horizon $n_\mathrm{p} = 16$, the RL policy does not yield further benefits and may even increase cost, suggesting limited generalisation beyond the training configuration.

Overall, these results indicate that RL-based adaptation can enhance performance for the configuration it is trained on, but

CC-MPC-RL	$n_\mathrm{p}=8$ $n_\mathrm{of}=100$	$n_\mathrm{p}=8$ $n_\mathrm{of}=200$	$n_\mathrm{p}=16$ $n_\mathrm{of}=100$	$n_\mathrm{p}=16$ $n_\mathrm{of}=200$
Total violation (mean) [%]	-4.2 %	-7.43 %	-41.88 %	-58.59 %
Total violation (variance) [%]	7.57 %	11.2 %	27.43 %	24.24 %
Violation count (mean) [%]	-2.2 %	-5.81 %	-35.82 %	-50.98 %
Violation count (variance) [%]	6.01 %	12.6 %	26.53 %	24.49 %
Cost (mean) [€]	10808.43 €	27051.14 €	387401.87 €	711028.66 €
Cost (variance) [€]	9913.18 €	18838.99 €	225910.49 €	222442.33 €

TABLE VI: CC-MPC-RL performance metrics across 50 simulations for varying prediction horizons (n_p) and offer set sizes (n_of), relative to the uncontrolled case.

that further work is needed to improve training stability and robustness across a broader range of operating conditions.

H. Comparison Between Methods

Tables VII–X summarise all approaches across the four configurations of prediction horizon ($n_\mathrm{p} \in \{8, 16\}$) and offer set size ($n_\mathrm{of} \in \{100, 200\}$), reported relative to the uncontrolled case.

$n_\mathrm{p}=8$ & $n_\mathrm{of}=100$	Greedy Matching	MPC	CC-MPC	CC-MPC-RL	MPC-PP
Total violation (mean) [%]	-0.26 %	-3.91 %	-3.93 %	-4.2 %	-4.54 %
Total violation (variance) [%]	0 %	5.97 %	7.56 %	7.57 %	7.78 %
Violation count (mean) [%]	0 %	-2.6 %	-1.88 %	-2.2 %	-1.94 %
Violation count (variance) [%]	0 %	7.42 %	6.02 %	6.01 %	6.25 %
Cost (mean) [€]	1501.40 €	16692.68 €	12258.2 €	10808.43 €	14042.17 €
Cost (variance) [€]	0 €	11060.39 €	10249.82 €	9913.18 €	9827.81 €

TABLE VII: Performance comparison of control approaches with respect to the uncontrolled case for $n_\mathrm{p}=8$, $n_\mathrm{of}=100$.

$n_\mathrm{p}=8$ & $n_\mathrm{of}=200$	Greedy Matching	MPC	CC-MPC	CC-MPC-RL	MPC-PP
Total violation (mean) [%]	-0.27 %	-6.57 %	-7.13 %	-7.43 %	-7.43 %
Total violation (variance) [%]	1.71 %	7.74 %	11.23 %	11.2 %	11.49 %
Violation count (mean) [%]	-0.32 %	-4.68 %	-5.74 %	-5.81 %	-5.0 %
Violation count (variance) [%]	0.77 %	9.58 %	12.64 %	12.6 %	10.77 %
Cost (mean) [€]	1944.92 €	22527.62 €	25623.13 €	27051.14 €	22273.09 €
Cost (variance) [€]	1181.43 €	12117.94 €	17058.65 €	18838.99 €	14293.88 €

TABLE VIII: Performance comparison of control approaches with respect to the uncontrolled case for $n_\mathrm{p}=8$, $n_\mathrm{of}=200$.

$n_\mathrm{p}=16$ & $n_\mathrm{of}=100$	Greedy Matching	MPC	CC-MPC	CC-MPC-RL	MPC-PP
Total violation (mean) [%]	-8.83 %	-35.81 %	-43.77%	-41.88 %	-41.49 %
Total violation (variance) [%]	11.89 %	25.0 %	27.51 %	27.43 %	27.46 %
Violation count (mean) [%]	-7.76 %	-29.22 %	-35.92 %	-35.82 %	-32.19 %
Violation count (variance) [%]	9.94 %	22.91 %	25.61 %	26.53 %	25.65 %
Cost (mean) [€]	102719.91 €	302572.72 €	371277.57 €	387401.87 €	313929.35 €
Cost (variance) [€]	59598.27 €	219901.91 €	230599.56 €	225910.49 €	230628.23 €

TABLE IX: Performance comparison of control approaches with respect to the uncontrolled case for $n_\mathrm{p}=16$, $n_\mathrm{of}=100$.

Three consistent trends emerge:

$n_\mathrm{p}=16$ & $n_\mathrm{of}=200$	Greedy Matching	MPC	CC-MPC	CC-MPC-RL	MPC-PP
Total violation (mean) [%]	-10.19 %	-56.18 %	-59.44 %	-58.59 %	-60.32 %
Total violation (variance) [%]	14.26 %	26.78 %	24.14 %	24.24 %	22.55 %
Violation count (mean) [%]	-5.53 %	-46.65 %	-50.85 %	-50.98 %	-48.8 %
Violation count (variance) [%]	6.46 %	27.55 %	22.51 %	24.49 %	21.13 %
Cost (mean) [€]	146396.78 €	538241.63 €	690521.43 €	711028.66 €	562637.84 €
Cost (variance) [€]	80776.80 €	241400.13 €	238974.97 €	222442.33 €	258304.89 €

TABLE X: Performance comparison of control approaches with respect to the uncontrolled case for $n_\mathrm{p}=16$, $n_\mathrm{of}=200$.

- **Baseline performance.** Across all configurations, Greedy Matching performs worst in terms of constraint violations, confirming that simple matching without predictive coordination is inadequate for structural CM.
- **Predictive versus probabilistic control.** Among predictive methods, MPC-PP achieves the lowest violations by exploiting perfect foresight, often at lower cost than the chance-constrained methods; however, recall that this is an ideal case, solely considered for a comparison. The CC-MPC and CC-MPC-RL attain nearly the same level of constraint satisfaction as MPC-PP while relying on the same imperfect forecasts as nominal MPC. This comes at substantially higher cost due to preventive activations of flexibility in anticipation of possible violations that may not materialise.
- **Effect of horizon and offer quantity.** Increasing the prediction horizon from $n_\mathrm{p}=8$ to $n_\mathrm{p}=16$ significantly improves performance for all predictive controllers. For example, the mean total violation of MPC decreases from -6.57% to -56.18% for $n_\mathrm{of}=200$. Similarly, larger offer sets reduce violations by providing more options to alleviate congestion. This underscores the importance of market design and ensuring that sufficient flexibility is available and properly structured in time.

I. Discussion and Summary

The numerical results demonstrate that predictive and probabilistic control strategies are highly effective for managing structural congestion in transmission networks. Uncontrolled operation and Greedy Matching fail to provide adequate safety, whereas all model-based approaches (MPC, MPC-PP, CC-MPC, and CC-MPC-RL) achieve substantial reductions in both the magnitude and frequency of line overloads.

The nominal MPC coordinates up- and down-regulation actions efficiently, even when driven by imperfect SARIMA-based forecasts. The MPC-PP benchmark shows that improved forecast accuracy simultaneously enhances both safety and economic efficiency. The chance-constrained controllers achieve constraint satisfaction levels close to those of MPC-PP while using the same imperfect forecasts as nominal MPC, at the cost of higher flexibility activation and increased operational cost.

Tuning the safety parameter α provides a direct mechanism for trading robustness against cost, although the effect is limited in the presence of structural congestion that cannot be fully eliminated. Increasing the number of offers and

extending the prediction horizon both enhance controllability by enlarging the pool of matchable flexibility, but also increase computational effort and may exhibit diminishing returns once sufficient flexibility is present.

Finally, the RL-enhanced CC-MPC shows that learning-based adaptation of uncertainty models can further improve performance for specific configurations, even when training is not fully converged. However, the lack of robust generalisation across horizons and offer sets points to the need for more systematic training strategies and improved RL formulations. Overall, the results confirm that combining predictive control, probabilistic constraint handling, and appropriately designed flexibility markets provides a powerful approach to CM under uncertainty, even when only approximate forecasts are available.

VII. Conclusion and future work

A combined CC-MPC–RL framework was developed to integrate probabilistic forecasting, model-based control, and learning-based adaptation for CM. The CC-MPC formulation expressed the CM problem as a dynamic stochastic optimisation, explicitly accounting for uncertainty in grid conditions while reflecting market rules from the Dutch CM framework. Real-world data from EDSN were used to construct statistical models of renewable generation variability, thereby grounding the uncertainty representation in realistic conditions. The reinforcement learning component provided adaptive tuning of model uncertainty parameters, enabling improved responsiveness and robustness under time-varying conditions.

Simulation results demonstrated that the proposed CC-MPC–RL approach effectively anticipated congestion, coordinated flexibility activation, and achieved superior constraint satisfaction compared to deterministic MPC and heuristic strategies. Incorporating probabilistic forecasts enhanced robustness against uncertainty, while the adaptive learning layer improved control performance over time. Together, these results confirm the feasibility and potential of a hybrid predictive–learning-based CM strategy in the Dutch context.

The framework bridges the gap between control theory and market-based CM. By embedding market mechanisms within an advanced control architecture. Overall, this study demonstrates that combining probabilistic forecasting, CC-MPC, and RL provides a promising and scalable pathway for data-driven, adaptive CM under uncertainty.T his integrated perspective not only highlights the potential of synergizing control and market paradigms but also opens up new research directions aimed at enhancing the robustness, applicability, and realism of the proposed framework. Several avenues for future research are proposed:

- **Enhanced uncertainty modelling:** Integrating ensemble forecasting, volatility models, or deep learning-based predictors could improve the representation of time-varying stochastic behaviour and lead to more reliable control decisions.
- **Model applicability:** Extending the controller to a non-linear AC formulation or reduced-order non-linear approximation would expand its applicability to medium- and low-voltage networks, enabling a unified approach across grid levels.
- **Market representation:** Incorporating richer market offer types, such as linked, block, or exclusive bids, which could enhance flexibility participation in CM.

=

VIII. Glossary

AC	Alternating Current
ARMA	AutoRegressive Moving-Average
CC-MPC	Chance-Constrained Model Predictive Control
CM	Congestion Management
DQN	Deep Q-Network
EDSN	Energie Data Services Nederland
FTO	Flex-Time Offer
GOPACS	Grid Operators Platform for AnCillary Services
MAPE	Mean Absolute Precentage Error
MPC	Model Predictive Control
PO	Profile Offer
RES	Renewable Energy Source
RL	Reinforcement Learning
RMSE	Root Mean Square Error
SARIMA	Seasonal AutoRegressive Integrated Moving-Average

Appendix

$$
\begin{aligned}
G_{\mathrm{pp}}(k) &= \begin{bmatrix} g_{\mathrm{pp}}^{(1,2),1}(k) & \dots & g_{\mathrm{pp}}^{(1,2),|\mathscr{N}|}(k) \\ \vdots & \ddots & \vdots \\ g_{\mathrm{pp}}^{|\mathscr{E}|,1}(k) & \dots & g_{\mathrm{pp}}^{|\mathscr{E}|,|\mathscr{N}|}(k) \end{bmatrix}, \\
G_{\mathrm{pq}}(k) &= \begin{bmatrix} g_{\mathrm{pq}}^{(1,2),1}(k) & \dots & g_{\mathrm{pq}}^{(1,2),|\mathscr{N}|}(k) \\ \vdots & \ddots & \vdots \\ g_{\mathrm{pq}}^{|\mathscr{E}|,1}(k) & \dots & g_{\mathrm{pq}}^{|\mathscr{E}|,|\mathscr{N}|}(k) \end{bmatrix} \\
G_{\mathrm{qp}}(k) &= \begin{bmatrix} g_{\mathrm{qp}}^{(1,2),1}(k) & \dots & g_{\mathrm{qp}}^{(1,2),|\mathscr{N}|}(k) \\ \vdots & \ddots & \vdots \\ g_{\mathrm{qp}}^{|\mathscr{E}|,1}(k) & \dots & g_{\mathrm{qp}}^{|\mathscr{E}|,|\mathscr{N}|}(k) \end{bmatrix} \\
G_{\mathrm{qq}}(k) &= \begin{bmatrix} g_{\mathrm{qq}}^{(1,2),1}(k) & \dots & g_{\mathrm{qq}}^{(1,2),|\mathscr{N}|}(k) \\ \vdots & \ddots & \vdots \\ g_{\mathrm{qq}}^{|\mathscr{E}|,1}(k) & \dots & g_{\mathrm{qq}}^{|\mathscr{E}|,|\mathscr{N}|}(k) \end{bmatrix}
\end{aligned}
\tag{36}
$$

References

[1] J. Rogelj, M. den Elzen, N. Höhne, T. Fransen, H. Fekete, H. Winkler, R. Schaeffer, F. Sha, K. Riahi, and M. Meinshausen, "Paris Agreement climate proposals need a boost to keep warming well below 2 °C," *Nature*, vol. 534, pp. 631–639, 2016.

[2] Y. Abdelilah, A. A. Báscones, V. Anatolitis, H. Bahar, P. Bojek, F. Briens, T. Criswell, J. Moorhouse, K. Veerakumar, and L. M. Martinez, "Renewables 2024," 2024.

[3] European Commission, "The future of European competitiveness: Part A: A competitiveness strategy for Europe," tech. rep., European Commission, 2025.
[4] S. Butorac, "Eu electricity grids," briefing, European Parliamentary Research Service (EPRS), European Parliament, 2025.
[5] M. Attar, S. Repo, A. Mutanen, J. Rinta-Luoma, T. Väre, and K. Kukk, "Market integration and TSO-DSO coordination for viable Market-based congestion management in power systems," *Applied Energy*, vol. 353, p. 16, 2024.
[6] European Commission. Joint Research Centre., *Redispatch and Congestion Management: Future Proofing the European Power Market.* LU: Publications Office, 2024.
[7] Tennet, "Grid diagram." https://www.tennet.eu/grid-diagram.
[8] A. Kumar, S. Srivastava, and S. Singh, "A zonal congestion management approach using ac transmission congestion distribution factors," *Electric Power Systems Research*, vol. 72, pp. 85–93, 2004.
[9] K. Purchala, L. Meeus, D. Van Dommelen, and R. Belmans, "Usefulness of DC power flow for active power flow analysis," in *IEEE Power Engineering Society General Meeting, 2005*, pp. 454–459, 2005.
[10] R. Hemmati, H. Saboori, and M. A. Jirdehi, "Stochastic planning and scheduling of energy storage systems for congestion management in electric power systems including renewable energy resources," *Energy*, vol. 133, pp. 380–387, 2017.
[11] A. Kumar, S. Srivastava, and S. Singh, "A zonal congestion management approach using real and reactive power rescheduling," *IEEE Transactions on Power Systems*, vol. 19, pp. 554–562, 2004.
[12] S. R. Khuntia and R. Smeets, "Unlocking flexibility for congestion management with redispatch bids: Requirements for a new bid structure," in *2024 20th International Conference on the European Energy Market (EEM)*, pp. 1–6, 2024.
[13] M. Altın, Ö. Göksu, R. Teodorescu, P. Rodriguez, B.-B. Jensen, and L. Helle, "Overview of Recent Grid Codes for Wind Power Integration," in *12th International Conference on Optimization of Electrical and Electronic Equipment*, 2010.
[14] B. Zarrouki, C. Wang, and J. Betz, "Adaptive Stochastic Nonlinear Model Predictive Control with Look-ahead Deep Reinforcement Learning for Autonomous Vehicle Motion Control," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 12726–12733, 2024.
[15] A. Tavakoli, F. Pardo, and P. Kormushev, "Action branching architectures for deep reinforcement learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2017.

Bibliography

[1] J. Rogelj, M. den Elzen, N. Höhne, T. Fransen, H. Fekete, H. Winkler, R. Schaeffer, F. Sha, K. Riahi, and M. Meinshausen, "Paris Agreement climate proposals need a boost to keep warming well below 2 °C," *Nature*, vol. 534, pp. 631–639, 2016.

[2] Y. Abdelilah, A. A. Báscones, V. Anatolitis, H. Bahar, P. Bojek, F. Briens, T. Criswell, J. Moorhouse, K. Veerakumar, and L. M. Martinez, "Renewables 2024," 2024.

[3] European Environment Agency., *Renewables, Electrification and Flexibility for a Competitive EU Energy System Transformation by 2030.* LU: Publications Office, 2025.

[4] European Commission, "The future of European competitiveness: Part A: A competitiveness strategy for Europe," tech. rep., European Commission, 2025.

[5] S. Butorac, "Eu electricity grids," briefing, European Parliamentary Research Service (EPRS), European Parliament, 2025.

[6] M. Attar, S. Repo, A. Mutanen, J. Rinta-Luoma, T. Väre, and K. Kukk, "Market integration and TSO-DSO coordination for viable Market-based congestion management in power systems," *Applied Energy*, vol. 353, p. 16, 2024.

[7] C. Zhao, E. Mallada, S. Low, and J. Bialek, "A unified framework for frequency control and congestion management," in *2016 Power Systems Computation Conference (PSCC)*, (Genoa, Italy), pp. 1–7, IEEE, 2016.

[8] European Commission. Joint Research Centre., *Redispatch and Congestion Management: Future Proofing the European Power Market.* LU: Publications Office, 2024.

[9] S. Peyghami, P. Davari, M. Fotuhi-Firuzabad, and F. Blaabjerg, "Standard Test Systems for Modern Power System Analysis: An Overview," *IEEE Industrial Electronics Magazine*, vol. 13, pp. 86–105, 2019.

[10] F. Tanrisever, K. Derinkuyu, and G. Jongen, "Organization and functioning of liberalized electricity markets: An overview of the Dutch market," *Renewable and Sustainable Energy Reviews*, vol. 51, pp. 1363–1374, 2015.

[11] S. R. Khuntia and R. Smeets, "Unlocking flexibility for congestion management with redispatch bids: Requirements for a new bid structure," in *2024 20th International Conference on the European Energy Market (EEM)*, pp. 1–6, 2024.

[12] Tennet, "Grid diagram." https://www.tennet.eu/grid-diagram.

[13] S. H. Low, "A three-phase power flow model and balanced network analysis," *arXiv preprint arXiv:2207.12519*, 2022.

[14] B. Sereeter, C. Vuik, C. Witteveen, and P. Palensky, "Optimal power flow formulations and their impacts on the performance of solution methods," in *2019 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, IEEE, 2019.

[15] S. Makridakis and M. Hibon, "ARMA Models and the Box–Jenkins Methodology," *Journal of Forecasting*, vol. 16, pp. 147–163, 1997.

[16] S. Jaggia, "Forecasting with ARMA Models," *Case Studies In Business, Industry And Government Statistics*, vol. 4, pp. 59–65, 2014.

[17] S. Shahriari, G. , Milad, S. , S. A., and T. and Rashidi, "Ensemble of ARIMA: Combining parametric and bootstrapping technique for traffic flow prediction," *Transportmetrica A: Transport Science*, vol. 16, pp. 1552–1573, 2020.

[18] R. Ningombam, C. Singh, S. Sreekumar, R. Bhakar, and S. Padmanaban, "Box–Cox integrated sARIMA model for day-ahead inertia forecasting," *Electrical Engineering*, vol. 107, pp. 9135–9153, 2025.

[19] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, pp. 461–464, 1978.

[20] H. Akaike, "Information Theory and an Extension of the Maximum Likelihood Principle," in *Proceedings of the 2nd International Symposium on Information Theory*, pp. 267–281, 1973.

[21] R. J. Hyndman and Y. Khandakar, "Automatic Time Series Forecasting: The **forecast** Package for *R*," *Journal of Statistical Software*, vol. 27, 2008.

[22] B. Kouvaritakis and M. Cannon, *Model Predictive Control*. Advanced Textbooks in Control and Signal Processing, Cham: Springer International Publishing, 2016.

[23] Schildbach, Georg, Fagiano, Lorenzo, and M. Morari, "Randomized Solutions to Convex Programs with Multiple Chance Constraints," *SIAM Journal on Optimization*, vol. 23, pp. 2479–2501, 2013.

[24] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro, "Sample Average Approximation Method for Chance Constrained Programming: Theory and Applications," *Journal of Optimization Theory and Applications*, vol. 142, pp. 399–416, 2009.

[25] G. Schildbach and M. Morari, "Scenario MPC for linear time-varying systems with individual chance constraints," in *2015 American Control Conference (ACC)*, pp. 415–421, 2015.

[26] A. B. Kordabad, D. Reinhardt, A. S. Anand, and S. Gros, "Reinforcement Learning for MPC: Fundamentals and Current Challenges," *IFAC-PapersOnLine*, vol. 56, pp. 5773–5780, 2023.

[27] K. Syranidis, M. Robinius, and D. Stolten, "Control techniques and the modeling of electrical power flow across transmission networks," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 3452–3467, 2018.

[28] S. Yang and Y. Zhu, "Distributed Stochastic ACOPF Based on Consensus ADMM and Scenario Reduction," in *2024 7th International Conference on Power and Energy Applications (ICPEA)*, pp. 604–609, 2024.

[29] K. Purchala, L. Meeus, D. Van Dommelen, and R. Belmans, "Usefulness of DC power flow for active power flow analysis," in *IEEE Power Engineering Society General Meeting, 2005*, pp. 454–459, 2005.

[30] M. Pantoš, "Market-based congestion management in electric power systems with increased share of natural gas dependent power plants," *Energy*, vol. 36, pp. 4244–4255, 2011.

[31] Y. Xu, H. Sun, H. Liu, and Q. Fu, "Distributed solution to DC optimal power flow with congestion management," *International Journal of Electrical Power & Energy Systems*, vol. 95, pp. 73–82, 2018.

[32] R. E. Bohn, M. C. Caramanis, and F. C. Schweppe, "Optimal Pricing in Electrical Networks over Space and Time," *The RAND Journal of Economics*, vol. 15, pp. 360–376, 1984.

[33] M. Jafarian, J. M. Scherpen, K. Loeff, M. Mulder, and M. Aiello, "A combined nodal and uniform pricing mechanism for congestion management in distribution power networks," *Electric Power Systems Research*, vol. 180, p. 106088, 2020.

[34] A. Jokic, P. P. J. van den Bosch, A. Virag, and R. M. Hermans, "On zonal pricing for congestion management," in *2012 9th International Conference on the European Energy Market*, p. 7, 2012.

[35] S. Gumpu, B. Pamulaparthy, and A. Sharma, "Review of Congestion Management Methods from Conventional to Smart Grid Scenario," *International Journal of Emerging Electric Power Systems*, vol. 20, 2019.

[36] M. A. Fotouhi Ghazvini, G. Lipari, M. Pau, F. Ponci, A. Monti, J. Soares, R. Castro, and Z. Vale, "Congestion management in active distribution networks through demand response implementation," *Sustainable Energy, Grids and Networks*, vol. 17, p. 13, 2019.

[37] I. Kalogeropoulos and H. Sarimveis, "Predictive control algorithms for congestion management in electric power distribution grids," *Applied Mathematical Modelling*, vol. 77, pp. 635–651, 2020.

[38] F. Shen, Q. Wu, S. Huang, X. Chen, H. Liu, and Y. Xu, "Two-tier demand response with flexible demand swap and transactive control for real-time congestion management in distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 114, p. 13, 2020.

[39] S. Huang and Q. Wu, "Real-Time Congestion Management in Distribution Networks by Flexible Demand Swap," *IEEE Transactions on Smart Grid*, vol. 9, pp. 4346–4355, 2018.

[40] B. Van Der Holst, G. Verhoeven, P. H. Nguyen, J. Morren, and K. Kok, "The activation of congestion service contracts for budget-constrained congestion management," *Electric Power Systems Research*, vol. 235, p. 7, 2024.

[41] R. Ciavarella, M. Di Somma, G. Graditi, and M. Valenti, "Congestion Management in distribution grid networks through active power control of flexible distributed energy resources," in *2019 IEEE Milan PowerTech*, pp. 1–6, 2019.

[42] A. Haque, M. T. Rahman, P. Nguyen, and F. Bliek, "Smart curtailment for congestion management in LV distribution network," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pp. 1–5, 2016.

[43] M. Esmaili, H. A. Shayanfar, and N. Amjady, "Multi-objective congestion management incorporating voltage and transient stabilities," *Energy*, vol. 34, pp. 1401–1412, 2009.

[44] R. Gupta, F. Sossan, and M. Paolone, "Grid-Aware Distributed Model Predictive Control of Heterogeneous Resources in a Distribution Network: Theory and Experimental Validation," *IEEE Transactions on Energy Conversion*, vol. 36, pp. 1392–1402, 2021.

[45] E. Luo, P. Cong, H. Lu, and Y. Li, "Two-Stage Hierarchical Congestion Management Method for Active Distribution Networks With Multi-Type Distributed Energy Resources," *IEEE Access*, vol. 8, p. 12, 2020.

[46] D.-T. Hoang, S. Olaru, A. Iovine, J. Maeght, P. Panciatici, and M. Ruiz, "Power Congestion Management of a sub-Transmission Area Power Network using Partial Renewable Power Curtailment via MPC," in *2021 60th IEEE Conference on Decision and Control (CDC)*, pp. 6351–6358, 2021.

[47] C. Straub, S. Olaru, J. Maeght, and P. Panciatici, "Zonal Congestion Management Mixing Large Battery Storage Systems and Generation Curtailment," in *2018 IEEE Conference on Control Technology and Applications (CCTA)*, pp. 988–995, 2018.

[48] K. Chakravarthi, P. Bhui, N. K. Sharma, and B. C. Pal, "Real Time Congestion Management Using Generation Re-Dispatch: Modeling and Controller Design," *IEEE Transactions on Power Systems*, vol. 38, pp. 2189–2203, 2023.

[49] N. Dkhili, S. Olaru, A. Iovine, M. Ruiz, J. Maeght, and P. Panciatici, "Predictive control based on stochastic disturbance trajectories for congestion management in subtransmission grids," *IFAC-PapersOnLine*, vol. 55, pp. 302–307, 2022.

[50] M. Siemonsmeier, M. von Heel, and A. Moser, "Impact of Uncertainties on Grid Congestion Management Measures with Long Lead Time," in *2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*, pp. 1–6, 2021.

[51] M. Ono, U. Topcu, M. Yo, and S. Adachi, "Risk-limiting power grid control with an ARMA-based prediction model," in *52nd IEEE Conference on Decision and Control*, pp. 4949–4956, 2013.

[52] M. Negnevitsky, P. Mandal, and A. K. Srivastava, "Machine Learning Applications for Load, Price and Wind Power Prediction in Power Systems," in *2009 15th International Conference on Intelligent System Applications to Power Systems*, pp. 1–6, 2009.

[53] D. Markovics and M. J. Mayer, "Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction," *Renewable and Sustainable Energy Reviews*, vol. 161, p. 112364, 2022.

[54] A. Hernandez-Matheus, K. Berg, V. Gadelha, M. Aragüés-Peñalba, E. Bullich-Massagué, and S. Galceran-Arellano, "Congestion forecast framework based on probabilistic power flow and machine learning for smart distribution grids," *International Journal of Electrical Power & Energy Systems*, vol. 156, p. 10, 2024.

[55] M. Hojjat and M. H. Javidi D. B., "Probabilistic Congestion Management Considering Power System Uncertainties Using Chance-constrained Programming," *Electric Power Components and Systems*, vol. 41, pp. 972–989, 2013.

[56] P. Omrani, H. Yektamoghadam, A. Nikoofard, M. R. Salehizadeh, and J. J. Liu, "Dynamic Congestion Management With Chance-Constrained MPC in Networked Microgrids Under Consumers-Related Uncertainties," *IEEE Transactions on Consumer Electronics*, p. 9, 2024.

[57] N. Dkhili, S. Olaru, A. Iovine, G. Giraud, J. Maeght, P. Panciatici, and M. Ruiz, "Data-Based Predictive Control for Power Congestion Management in Subtransmission Grids Under Uncertainty," *IEEE Transactions on Control Systems Technology*, vol. 31, pp. 2146–2158, 2023.

[58] A. J. Conejo and X. Wu, "Robust optimization in power systems: A tutorial overview," *Optimization and Engineering*, vol. 23, pp. 2051–2073, 2022.

[59] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 2nd ed., 2014.

[60] H. Musbah and M. El-Hawary, "SARIMA Model Forecasting of Short-Term Electrical Load Data Augmented by Fast Fourier Transform Seasonality Detection," in *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, pp. 1–4, 2019.

[61] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *European Transport Research Review*, vol. 7, no. 3, p. 21, 2015.

[62] K. Fukuda, "Time-Series Forecast Jointly Allowing the Unit-Root Detection and the Box–Cox Transformation," *Communications in Statistics - Simulation and Computation*, vol. 35, no. 2, pp. 419–427, 2006.

[63] M. G. Pinheiro, S. C. Madeira, and A. P. Francisco, "Short-term electricity load forecasting—A systematic approach from system level to secondary substations," *Applied Energy*, vol. 332, p. 120493, 2023.

[64] S. Fan and R. J. Hyndman, "Short-Term Load Forecasting Based on a Semi-Parametric Additive Model," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 134–141, 2012.

[65] A. Clements, A. Hurn, and Z. Li, "Forecasting day-ahead electricity load using a multiple equation time series approach," *European Journal of Operational Research*, vol. 251, no. 2, pp. 522–530, 2016.

[66] A. Kumar, S. Srivastava, and S. Singh, "A zonal congestion management approach using ac transmission congestion distribution factors," *Electric Power Systems Research*, vol. 72, pp. 85–93, 2004.

[67] R. Hemmati, H. Saboori, and M. A. Jirdehi, "Stochastic planning and scheduling of energy storage systems for congestion management in electric power systems including renewable energy resources," *Energy*, vol. 133, pp. 380–387, 2017.

[68] A. Kumar, S. Srivastava, and S. Singh, "A zonal congestion management approach using real and reactive power rescheduling," *IEEE Transactions on Power Systems*, vol. 19, pp. 554–562, 2004.

[69] A. Bemporad and M. Morari, "Control of systems integrating logic, dynamics, and constraints," *Automatica*, vol. 35, no. 3, pp. 407–427, 1999.

[70] J. Hu, Y. Shan, Y. Yang, A. Parisio, Y. Li, N. Amjady, S. Islam, K. W. Cheng, J. M. Guerrero, and J. Rodríguez, "Economic Model Predictive Control for Microgrid Optimization: A Review," *IEEE Transactions on Smart Grid*, vol. 15, pp. 472–484, 2024.

[71] M. Altın, Ö. Göksu, R. Teodorescu, P. Rodriguez, B.-B. Jensen, and L. Helle, "Overview of Recent Grid Codes for Wind Power Integration," in *12th International Conference on Optimization of Electrical and Electronic Equipment*, 2010.

[72] B. Zarrouki, C. Wang, and J. Betz, "Adaptive Stochastic Nonlinear Model Predictive Control with Look-ahead Deep Reinforcement Learning for Autonomous Vehicle Motion Control," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 12726–12733, 2024.

[73] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with Deep Reinforcement Learning," 2013.

[74] H. Van Hasselt, A. Guez, and D. Silver, "Deep Reinforcement Learning with Double Q-Learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, 2016.

[75] A. Tavakoli, F. Pardo, and P. Kormushev, "Action branching architectures for deep reinforcement learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2017.

[76] Z. Wang, T. Schaul, M. Hessel, H. Van Hasselt, M. Lanctot, and N. De Freitas, "Dueling network architectures for deep reinforcement learning," in *Proceedings of the 33rd International Conference on International Conference on Machine*, vol. 48, p. 1995–2003, 2016.

Glossary

List of Acronyms

AC	Alternating Current
ACF	Autocorrelation Function
ARMA	AutoRegressive Moving-Average
ARIMA	AutoRegressive Integrated Moving-Average
CC-MPC	Chance-Constrained Model Predictive Control
CM	Congestion Management
DC	Direct Current
DQN	Deep Q-Network
EDSN	Energie Data Services Nederland
FTO	Flex-Time Offer
GOPACS	Grid Operators Platform for AnCillary Services
MAPE	Mean Absolute Precentage Error
MPC	Model Predictive Control
PACF	Partial Autocorrelation Function
PO	Profile Offer
RES	Renewable Energy Source
RL	Reinforcement Learning
RMSE	Root Mean Square Error
SARIMA	Seasonal AutoRegressive Integrated Moving-Average