Universidad de Concepción
Facultad de Ingeniería

REAL-TIME DEMAND RESPONSE SCHEDULING BASED ON CUSTOMIZED ONLINE LEARNING OF CUSTOMERS CONSIDERING UNCERTAINTY

Tesis presentada a la Facultad de Ingeniería de la Universidad de Concepción para optar al grado académico de

Doctor en Ciencias de la Ingeniería con Mención en Ingeniería Eléctrica

POR: LESTER JULIO MARRERO RODRÍGUEZ

Profesor Guía: Ph.D. Daniel Sbárbaro Hofer
Profesor Co-guía: Dr. Luis García Santander

Concepción, Chile, abril de 2024

Dedicado a mi familia.

Abstract

Chilean commercial, public, and residential sector consume 35 % of the electrical energy in the country. Given the continuous economic growth of society, this percentage will increase gradually, as expected. However, Chile has also made efforts to transition towards power decarbonization, in line with the global context that seeks to reduce its carbon footprint. Distribution grids are a complex infrastructure and, within this scenario, are reaching an intelligent configuration. As a result of this transformation, the proportion of renewable energy sources (RESs) is constantly increasing, and advanced metering systems are playing an essential role in recording electricity data to provide insights into customers' behavior and corresponding lifestyles. Conditioned by the gradual deregulation of the distribution sector, new market agents are participating in the electricity market, making customers' service more active and competitive. Therefore, integrating renewable capacity by exploiting the operational flexibility that can arrive from scheduling demand is vital in the current distribution system. Efficient scheduling can also help to decouple economic growth from energy consumption. However, both renewable production and demand exhibit a stochastic behavior, which leads to an increase in uncertainty due to forecast errors in generation and deviations in demand in real-time. This uncertainty jeopardizes the safe operation of the distribution system.

This thesis focuses on real-time demand response (DR) scheduling based on customized online learning of customers' behavior considering uncertainty.

First, the thesis proposes an online framework to characterize demand response (DR) over time. The approach facilitates obtaining and updating the daily consumption patterns of customers. The essential concept of response profile class (RPC) is introduced for characterization, complemented by the measure of the variability in customer behavior. For daily profiles, the work uses a modified version of the incremental clustering by fast search and find of density peaks (CFSFDP) algorithm

that considers the multivariate normal kernel density estimator and incremental forms of the Davies-Bouldin and Xie-Beni validity indices. Case studies conducted using real-world and simulated daily profiles of residential and commercial Chilean end-users demonstrate how the proposed approach can continuously characterize DR. Results prove that the presented framework achieves realistic customer models for effective energy management by estimating customer response to the price signal at the distribution system operator (DSO) level.

Second, the thesis proposes an online framework for scheduling customers' power responses to support integrating photovoltaic (PV) generation into the distribution system. The approach helps the management of demand flexibilities by providing optimal control prices in real-time. A bi-level optimization model is used to model interactions between the DSO in the upper level, which seeks to maximize the profit, and customers in the lower level, who reduce their electricity bills. In addition, to obtain the RPCs of customers, a characterization stage is introduced by applying the CFSFDP algorithm to daily profiles. The pricing problem, however, is highly challenging since the DSO also requires ensuring the reliable operation of the distribution grid and dealing with uncertainties in consumption and renewable production. Accordingly, the work includes and reformulates chance constraints (CCs) for squared nodal voltage and complex power flow in lines. The chance-constrained bi-level problem converts to an equivalent mixed-integer second-order cone programming problem, embedded into a model predictive control to exploit newly available information of the system states. A case study using real-world local market prices and daily profiles of residential and commercial Chilean end-users on the IEEE-37 node test distribution feeder demonstrates how the proposed framework can schedule DR considering uncertainty.

Third, for optimal scheduling of other distribution-level energy resources, the previous framework is expanded by including dispatchable inverters of PV facilities. The approach also addresses uncertainty in the distribution system modeling by incorporating, in addition to the above, CCs for the apparent power of PV inverters. A case study with real-world local market prices and daily profiles of Chilean residential and commercial end-users on the IEEE-37 node test distribution feeder demonstrates how the presented framework enables optimal real-time scheduling of customers and PV facilities.

The investigation has significant implications both technically, by enhancing the efficiency and reliability of the distribution system, and economically, by generating financial benefits for market players. Furthermore, its methodology is applicable and suitable to meet the practical requirements of Chilean society.

Resumen

En Chile, el 35 % del consumo de energía eléctrica se debe al sector comercial, público y residencial. Es esperado que este valor aumente sucesivamente dado el continuo desarrollo económico de la sociedad. Por otro lado, el país se encuentra desde hace algunos años inmerso en un proceso de transición marcado por un contexto internacional que busca la descarbonización de la matriz energética, para disminuir con ello la huella de carbono. Los sistemas eléctricos, y en particular las redes de distribución constituyen una infraestructura muy compleja, y en medio de esta panorámica están adquiriendo una configuración de red inteligente. En la medida de esta transformación, la proporción de fuentes de energía renovable, esencial para el futuro energético sostenible, crece constantemente. Los sistemas de medición inteligente, además, están desempeñando un papel vital al registrar datos de consumo que proporcionan información sobre el comportamiento y estilos de vida de los clientes. Condicionado por la gradual desregulación del segmento de distribución, nuevos actores toman parte también en el mercado de electricidad y el servicio al cliente se hace más activo y competitivo. Por consiguiente, la integración de la creciente capacidad renovable mediante la explotación de la flexibilidad de operación que puede obtenerse a partir de la programación de la demanda resulta un pilar básico en la red de distribución actual. Una programación eficiente garantiza además el desacople entre el crecimiento económico y el consumo de energía. Sin embargo, una propiedad inherente tanto de la generación renovable como de la demanda es su comportamiento estocástico. Por tanto, el aumento de la incertidumbre a partir de errores de pronóstico en la generación y desviaciones de la demanda en tiempo real compromete la operación segura del sistema.

Esta tesis tiene como objetivo la programación de la respuesta de la demanda (DR) en tiempo real a partir del aprendizaje en línea personalizado de los clientes considerando la incertidumbre.

La primera parte de la tesis propone una metodología en línea para caracterizar en el tiempo la DR. El enfoque facilita la obtención y actualización de los patrones diarios de consumo de los clientes. Para la caracterización, se introduce el concepto esencial de clase de perfil de respuesta (RPC), complementado con la medida de la variabilidad en el comportamiento del cliente. El trabajo utiliza para los perfiles diarios una versión modificada del algoritmo incremental *clustering by fast search and find of density peaks* (CFSFDP) que considera el estimador de densidad *kernel* normal multivariante y formas incrementales de los índices de validación Davies-Bouldin y Xie-Beni. Se presentan estudios de caso que demuestran cómo el enfoque propuesto permite caracterizar continuamente la DR empleando perfiles diarios reales y simulados de clientes residenciales y comerciales del sur de Chile. Los resultados prueban que la metodología logra modelos de clientes realistas para una gestión eficaz de la energía al estimar la respuesta de estos a una señal de precio a nivel del operador del sistema de distribución (DSO).

La segunda parte de la tesis propone una metodología en línea para programar las respuestas de demanda de los clientes y favorecer la integración de la generación fotovoltaica en el sistema de distribución. El enfoque propuesto contribuye a la gestión de la flexibilidad de la demanda al proporcionar precios de control óptimos en tiempo real. Se utiliza un modelo de optimización de dos niveles para modelar las interacciones entre el DSO en el nivel superior, que persigue maximizar su ganancia, y los clientes en el nivel inferior, que buscan reducir sus facturas de electricidad. Para obtener las RPC de los clientes, se introduce una etapa de caracterización aplicando el algoritmo CFSFDP a los perfiles diarios. Sin embargo, la generación de precios de control plantea un gran desafío, ya que el DSO requiere garantizar también la operación confiable de la red y abordar la incertidumbre en el consumo y la producción renovable. En consecuencia, el trabajo incluye y reformula restricciones probabilísticas para el voltaje cuadrático en nodos y el flujo de potencia complejo en líneas. El problema de dos niveles con restricciones probabilísticas se transforma en un problema equivalente de programación de cono de segundo orden entera mixta, embebido en un modelo de control predictivo para explotar información disponible de los estados del sistema. Un estudio de caso que emplea precios locales y perfiles diarios de los clientes chilenos residenciales y comerciales en el alimentador de prueba de 37 nodos de IEEE demuestra cómo el enfoque permite programar la DR en tiempo real, considerando la incertidumbre en el voltaje en nodos y el flujo de potencia en líneas.

En la tercera parte de la tesis, para abordar la programación óptima de otros recursos de energía a nivel de sistema de distribución, se expande el análisis de la segunda parte

considerando el control de los inversores de las instalaciones fotovoltaicas. Un nuevo estudio de caso con precios locales y perfiles diarios de los clientes chilenos residenciales y comerciales en el alimentador de prueba de 37 nodos de IEEE demuestra cómo el enfoque presentado permite la programación óptima en tiempo real tanto de los clientes como de las instalaciones fotovoltaicas, incluyendo, además de la incertidumbre previa, la correspondiente a la potencia aparente de los inversores.

La investigación desarrollada tiene gran importancia tanto desde el punto de vista técnico, al propiciar una operación más eficiente y confiable del sistema de distribución, como económico, con ganancias financieras para los participantes del mercado. La misma es aplicable según su metodología y se ajusta a la exigencia práctica para la sociedad chilena.

Contents

Nomenclature

General notation

In this thesis, (column) vectors are in bold lower case, such as $\mathbf{x}$; sets and matrices are interchangeably in bold upper case, such as $\mathbf{X}$. The notation $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that $\mathbf{x}$ is a normal random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Last, $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts of a complex number, the operator $\mathbb{E}[\cdot]$ indicates the expected value, and $[\cdot]^\top$ is used for transposition, $(\cdot)^*$ for the complex conjugate, and $\perp$ for the complementarity.

Sets, parameters, and variables

$\mathbf{L}$	Set of customers $\{1, \dots, L\}$
$\mathbf{T}$	Set of time points $\{1, \dots, T\}$
$\bar{\mathbf{p}}_l$	Expected active power profile of customer l $[\bar{p}_{l_1} \ \dots \ \bar{p}_{l_T}]^\top$
$\bar{p}_{l_t}$	Expected active power of customer l at time point t
$\bar{q}_{l_t}$	Expected reactive power of customer l at time point t
$\bar{s}_{l_t}$	Expected complex power of customer l at time point t
$p_{l_t}^{\min}, p_{l_t}^{\max}$	Minimum and maximum power bounds for $\bar{p}_{l_t}$
$r_{l_t}^{\mathrm{d}}, r_{l_t}^{\mathrm{u}}$	Maximum ramp-down rate and ramp-up rate for $\bar{p}_{l_t}$
e_l	Minimum daily energy for customer l
Δt	Interval between two consecutive time points
τ	Initial number of days
$\mathbf{p}_i^0$	ith initial active power profile $[p_{i\,1}^0 \ \dots \ p_{i\,T}^0]^\top$
$\mathbf{P}_0$	Initial set of daily profiles $\{\mathbf{p}_i^0, i = 1, \dots, N_0\}$
$\mathbf{p}_l$	Active power profile of customer l $[p_{l_1} \ \dots \ p_{l_T}]^\top$
$\mathbf{P}$	Set of daily profiles $\{\mathbf{p}_l\}$

ρ	Local density
δ	Minimum distance
γ	Product of ρ and δ
h_t	Smoothing parameter of time point t
$\mathbf{c}_k^0, \mathbf{c}_k$	Cluster center of cluster k, $k = 1, \dots, K$
$\mathbf{G}_k^0, \mathbf{G}_k$	Cluster k of daily profiles $\left[\mathbf{c}_k^0 \; \mathbf{p}_{k,2}^0 \; \cdots \; \mathbf{p}_{k,N_k^0}^0 \right]$, $\left[\mathbf{c}_k^0 \; \mathbf{p}_{k,2} \; \cdots \; \mathbf{p}_{k,N_k} \right]$
$s(\mathbf{G}_k^0), s(\mathbf{G}_k)$	Structure distance of $\mathbf{G}_k^0, \mathbf{G}_k$
N_r	Number of representative points
α	Shrink factor
$\mathbf{R}_k^0, \mathbf{R}_k$	Set of representative points of cluster k $\{ \mathbf{r}_{k,u}^0, u = 1, \dots, N_r \}$, $\{ \mathbf{r}_{k,u}, u = 1, \dots, N_r \}$
$\mathbf{A}$	Auxiliary matrix $\left[\mathbf{p}_{k,l} \; \mathbf{r}_{k,v}^0 \right]$
H_l	Variability of customer l
λ_{l_t}	Electricity price of customer l at time point t
$\beta_{l_t}^{\min}, \beta_{l_t}^{\max}$	Dual variables of $p_{l_t}^{\min}, p_{l_t}^{\max}$
$\psi_{l_t}^{\mathrm{d}}, \psi_{l_t}^{\mathrm{u}}$	Dual variables of $r_{l_t}^{\mathrm{d}}, r_{l_t}^{\mathrm{u}}$
ε_l	Dual variable of e_l
$\mathbf{G}$	Set of PV facilities $\{1, \dots, G\}$
p_{l_t}	Active power of customer l at time point t
q_{l_t}	Reactive power of customer l at time point t
$\bar{p}_{g_t}$	Active power forecast of PV facility g at time point t
$\bar{q}_{g_t}$	Reactive power forecast of PV facility g at time point t
$\bar{s}_{g_t}$	Complex power forecast of PV facility g at time point t
$\bar{p}_{g_t}^{\mathrm{av}}$	Forecast of available active power of PV facility g at time point t
p_{g_t}	Active power of PV facility g at time point t
q_{g_t}	Reactive power of PV facility g at time point t
$\mathbf{N}$	Set of nodes of the distribution system $\{1, \dots, N\}$
$\{0\}$	Slack node
$\mathbf{E}$	Set of distribution lines $\{(m, n) : m \in \mathbf{N} \cup \{0\}, n \in \mathbf{N}\}$
P_{n_t}	Active power of node n at time point t
Q_{n_t}	Reactive power of node n at time point t
S_{n_t}	Complex power of node n at time point t
v_{n_t}	Voltage magnitude of node n at time point t
V_{n_t}	Squared voltage of node n at time point t
$\bar{V}_{n_t}$	Squared voltage of node n at time point t with expected and forecast values
$V^{\min}, V^{\max}$	Minimum and maximum voltage limits for V_{n_t}

R_{mn}	Resistance of line mn
X_{mn}	Reactance of line mn
Z_{mn}	Impedance of line mn
P_{mn_t}	Active power flow from node m to node n at time point t
$\bar{P}_{mn_t}$	Active power flow from node m to node n at time point t with expected and forecast values
Q_{mn_t}	Reactive power flow from node m to node n at time point t
$\bar{Q}_{mn_t}$	Reactive power flow from node m to node n at time point t with expected and forecast values
S_{mn_t}	Complex power flow from node m to node n at time point t
$S_{mn}^{\max}$	Power flow limit of line mn
$\mathbf{p^n}_t$	Active power of nodes $[P_{1_t} \dots P_{N_t}]^{\top}$
$\mathbf{q^n}_t$	Reactive power of nodes $[Q_{1_t} \dots Q_{N_t}]^{\top}$
$\mathbf{D}$	$|\mathbf{E}| \times N$ binary matrix for mapping values of active and reactive power in nodes into power flows and voltages
$\mathbf{d}_{mn}$	Vector corresponding to the mnth row of $\mathbf{D}$
$\mathbf{d}_n$	Vector corresponding to the nth column of $\mathbf{D}$
$\mathbf{H}$	$[L + G] \times N$ binary matrix indicating the belonging of each customer l and PV facility g to the corresponding node n
$\mathbf{h}_n$	Vector corresponding to the nth column of $\mathbf{H}$
$\mathbf{p}_t$	Active power of customers and PV facilities at time point t
$\mathbf{q}_t$	Reactive power of customers and PV facilities at time point t
ξ_{l_t}	Deviation of the customer response
ξ_{g_t}	PV forecast error
$\boldsymbol{\xi}_t$	Vector of active power deviation or forecast error in the nodes
σ_{g_t}	Standard deviation of forecast errors for PV facility g at time point t
$\boldsymbol{\Theta}_t$	$N \times N$ diagonal matrix of tangents correspondingly relating the active and reactive nodal deviations or errors at time point t
$\mathbf{R}$	$N \times N$ diagonal matrix of line resistances
$\mathbf{X}$	$N \times N$ diagonal matrix of line reactances
λ_t^b	Locational marginal price at time point t
$\lambda^{\min}, \lambda^{\max}$	Minimum and maximum price bounds for λ_{l_t}
λ^r	Regulated price
ϵ_{v}	Violation probability for nodal voltages
ϵ_{i}	Violation probability for power flows in lines
ϵ_{g}	Violation probability for PV facilities
Φ	Cumulative distribution function of the standard normal distribution
$\mathbf{I}$	Set of sides $\{1, \dots, I\}$ of the inscribed polygon

W	$I \times 3$ matrix of coefficients
s	Vector of elements in $\mathbb{R}^3$
η_i	Slope of side i of the inscribed polygon
ξ_{1_i}, ξ_{2_i}	Vertices of side i on the inscribed polygon
θ	Angle between two consecutive vertices of the inscribed polygon
M^p, M^d	Positive constants denoting upper bounds for the primal and dual variables
$B_{l_t}^{\min}, B_{l_t}^{\max}$	Binary variables
$\Psi_{l_t}^{d}, \Psi_{l_t}^{u}$	Binary variables
E_l	Binary variable

Glossary

CC   Chance constraint
CDF   Cumulative distribution function
CFSFDP   Clustering by fast search and find of density peaks
DER   Distributed energy resource
DR   Demand response
DSO   Distribution system operator
ESS   Energy storage system
iDB   Incremental Davies-Bouldin index
iXB   Incremental Xie-Beni index
KKT   Karush-Kuhn-Tucker
LMP   Locational marginal price
MILP   Mixed-integer linear programming
MISOCP   Mixed-integer second-order cone programming
MPC   Model predictive control
OPF   Optimal power flow
PCA   Principal component analysis
PV   Photovoltaic
RES   Renewable energy source
RPC   Response profile class
SOC   Second-order cone

List of Figures

List of Tables

Chapter 1

Introduction

In 2022, disruptions in supply, rising energy prices, and record deployment and investment in renewables marked the energy sector. Solar photovoltaics (PV) accounted for 70 % of the total capacity additions of renewable power, where centralized utility-scale PV, for example, reached 124.8 GW of new installations driven by tenders and the attractiveness of power purchase agreements, and distributed PV added 115.2 GW.

As fundamental changes reshape the global energy system, policymakers, grid operators, and investors are becoming more aware of the role that renewable energy sources (RESs) play beyond mitigating climate change. Among the key benefits that renewables bring to a new energy system are energy security, economic and social value creation, and, potentially, higher geopolitical stability. However, integrating and expanding renewables to achieve high shares in utility power grids remains a significant challenge. Effective grid integration requires technology development, supporting policies, increased investment, and coordination among stakeholder groups. Also, their variable nature can affect system stability, making it more difficult to balance supply and demand, especially considering that for electricity demand, an estimated increase of 2 to 3 % is expected annually from 2021 to 2030 [1].

Technically, increasing shares of renewable power in power systems requires higher degrees of flexibility, not only from the electricity suppliers. Indeed, flexibility on the generation side usually leads to operating conventional units at production levels higher or lower than their optimum to accommodate the inherent variability of renewable generation by ramping up or down. Further, these ramping excursions may often end with the start-up or shutdown of conventional units. Consequently, high

variations of renewable generation to accommodate may result in power plants operating less efficiently, thus reducing the emissions savings brought about by the contribution of renewables to the electricity supply. However, a power system equipped with demand response (DR) can better cope with this by taking advantage of the flexibility of consumers [2]. In particular, the implementation of DR targets the control of the customer's power-consuming behavior to meet the following objectives: 1) reduction of the peak power consumption; 2) reduction of the total needed power generation, as the main result of the prior objective; 3) change of the demand to follow the available supply, especially with high penetration of RESs; and 4) elimination of overloads in the distribution system [3]. Time-varying pricing (also called dynamic pricing) can induce the DR of consumers, thereby improving economic efficiency and enhancing welfare concerning other forms [4].

As advanced metering infrastructures become more affordable, accessible, and spread in the distribution grid, more entities will take advantage of distributed energy resources (DERs), such as distributed PV, energy storage systems (ESSs), and DR. Thus, conventional passive customers will become active participants in the retail electricity market by providing DERs more efficiently and effectively [5].

1.1 Background

Integrating renewable capacity by exploiting the operational flexibility that can arrive from scheduling demand is vital in the current distribution system. Efficient scheduling also guarantees the decoupling between economic growth and electrical energy consumption, materializing in more environmental and financial benefits. The proposed approach in this thesis considers the management of demand flexibility by providing optimal control prices in real-time.

A critical challenge in the optimal scheduling of the distribution grid operation is that great uncertainty characterizes the required decisions, particularly with the increasing integration of weather-dependent generation facilities, which makes the available production capacity uncertain. Moreover, DERs that are being increasingly adopted close to individual consumption significantly increase the load uncertainty from the point of view of the distribution system operator (DSO). Thus, future distribution systems will involve great uncertainty on both the production and consumption sides [6].

Solving the scheduling problem on radial distribution grids also involves distribution systems' constraints relative to nodal voltages and power flows in lines. Traditionally, energy management approaches solve this problem as a deterministic programming model that targets a day-ahead scheduling task based on time points forecast for the

above variables. However, obtaining accurate forecasts of these uncertain variables is typically unattainable. Therefore, the increase in uncertainty of forecast errors in generation and deviations of demand in real-time can jeopardize the reliable operation of the distribution system in providing electricity to customers. In addition, the higher presence of local markets leads to an increase in intra-day electricity trading, thus distancing real-time from day-ahead scheduling.

Accounting for uncertainty in the scheduling of demand-side resources has been proposed in the literature by several stochastic methods, including robust optimization [6], stochastic programming [7], and chance-constrained formulations [8], [9]. The main difference between these approaches relies on how they address the uncertainty. Robust optimization, for instance, uses sets or distribution bounds to characterize the range of variability of the uncertain parameters [6]. In stochastic programming, uncertainty is described based on stochastic processes, conveniently characterized using scenarios [7]. In chance-constrained optimization problems, constraints state that the probability of a random event is kept smaller than a target value [8].

For several years now, a set of initiatives has been deployed in Chile to promote adequate legal frameworks for the continuous development of the country's distribution sector.

In the National Congress, the bill proposing changes to the right of customers' electrical portability, known as the Electrical Portability Bill [10], continues to be debated. The bill aims to modernize the regulatory framework of the distribution segment, building on the progress made in law No. 21.194 [11], as a joint effort of the Ministry of Energy and the National Energy Commission.

This bill is structured based on the following four main ideas: 1) grant all customers the right to choose their energy supplier, which allows them to have higher management over their energy consumption and potentially save money on their bills; 2) make supply tenders flexible, allowing for supplier choice and ensuring secure supply, compatible with recognizing the retailer as new market agent; 3) ensure the entry of new players to the electricity market, such as the Information Manager (this private and tendered body aims to guarantee complete independence of information management, protect end-users data, provide controlled and symmetrical access to interested parties, and offer tools to end-users to facilitate their decision-making process); and 4) respect the current tendered supply contracts whereas gradually transitioning to new ones.

Both organizations also seek to implement a flexibility strategy for the National Electrical System (SEN) [12]. The objective is to define actions that can help achieve

market signals and processes to facilitate the development and use of the necessary flexibility for the safe, efficient, and sustainable operation of the SEN. The strategy has various supporting features, namely the commitment to achieve carbon neutrality by 2050, the present and future rise in renewable generation influencing the system operation, the transformation in the electrical system's dynamics concerning energy flows, and the existence of significant net demand ramps with a predicted increase in future scenarios.

In defining the guidelines for future modifications of the regulatory framework in facing the above challenges, the strategy outlines the following actions: 1) developing a market design that fosters a more flexible power system; 2) creating a regulatory framework for ESSs and other emerging flexible technologies; and 3) ensuring the flexible operation of the power system. For example, the National Congress has recently passed Law No. 21.505 [13], published in 2022, to promote the introduction of storage technologies to facilitate the integration of RESs into the electrical mix and the use of electromobility in the distribution grid.

Summarizing, in recent years, the DR scheduling problem has been the focus of several solutions in the technical literature. However, most of these solutions do not assess the effect of the final consumption profiles on the distribution grid, and they also concentrate on the day-ahead market, leading to higher uncertainty in decision-making due to the separation of actual renewable generation from forecasted values or unexpected behaviors in load and electricity prices. This situation can create a complex scenario for the DSO because of the unavailability of fully dispatching some renewable resources in real-time (with the consequent power curtailments) or, conversely, the requirement of a greater reserve from the power system.

On the other hand, in the literature, the continuous characterization of customers' consumption behavior to understand their preferences and provide customized services in terms of DR has not yet been explored deeply. This study is essential for two reasons: 1) to accurately balance these preferences with the needs of the distribution operator; and 2) because each consumer's installed power and flexibility to respond to the price signal are not the same.

1.2 Hypothesis and Objective

The hypothesis of the thesis is as follows:

Using a stochastic optimization based on the continuous learning of the sensitivity of groups of customers to electricity price signals and their characterization contributes to the effective scheduling of DR in real-time.

The hypothesis is founded on the following scientific questions intended for the development of the thesis:

1) Can a price-based DR program be customized for groups of customers over time?

2) Is it feasible to determine in real-time the price signals that provide desired power responses from customers?

3) How can a management framework be defined to minimize operational costs and maintain the power balance in real time while considering grid conditions and uncertainties in demand and renewable generation?

Then, the objective of the thesis is as follows:

To develop a real-time DR scheduling based on customized online learning of customers considering uncertainty.

1.3   Contributions

The contributions of the thesis in addressing the above challenges are summarized below, divided into online DR characterization, online DR scheduling, and online scheduling of distribution-level energy resources.

1.2.1 Online DR Characterization

1) An innovative framework for online DR characterization is proposed. The thesis presents the modified incremental clustering by fast search and find of density peaks (CFSFDP) algorithm, defined to work in a Hilbert space. Online clustering introduces the multivariate normal kernel density estimator for robustness to the number of objects in clusters and the monitoring of algorithmic performance through incremental forms of the Davies-Bouldin (iDB) and Xie-Beni (iXB) validity indices.

2) Application of the proposed approach allows the DSO to perform two essential activities: updating the response profile class (RPC) and variability when customer response materializes (at the end of the day) and estimating customer behavior to price signals based on a known RPC (within the next day).

3) The framework is tested with real-world and simulated daily profiles. Results show the online process for obtaining RPCs of residential and commercial Chilean end-users. The thesis also provides a comparison and sensitivity analysis employing different combinations for the number of representatives and shrink factor.

1.2.2 Online DR Scheduling

1) An online framework for scheduling customers' power responses to integrate PV production into the distribution system is proposed. The thesis introduces a bi-level

optimization model where the DSO in the upper level pursues maximizing its profit by determining the optimal price signals, while customers in the lower level aim to reduce their electricity bills.

2) The proposed framework ensures the reliable operation of the distribution grid by limiting squared nodal voltages and complex power flows in lines with high probability. Specifically, chance constraints (CCs) are considered for these magnitudes and reformulated accordingly. The chance-constrained bi-level model then converts to an equivalent mixed-integer second-order cone programming (MISOCP) problem to find the globally optimal solution.

3) The framework is tested with the real-world daily profiles of residential and commercial Chilean end-users on the IEEE-37 node test distribution feeder. Results demonstrate how the proposed model can schedule DR in real time. The thesis also explores the impact of uncertainty on system operation and presents a comparison analysis with the deterministic formulation involving the DSO's profit and feasibility of the CCs.

1.2.3 Online Scheduling of Distribution-Level Energy Resources

1) An extension of the chance-constrained bi-level model to also determine optimal power set points for dispatchable inverters of PV facilities is proposed such that limits of squared nodal voltages, complex power flows in lines, and inverters' capacity are satisfied with high probability.

2) The framework is tested in a new case study with the real-world daily profiles of Chilean residential and commercial end-users on the IEEE-37 node test distribution feeder. Results demonstrate how the presented framework enables optimal real-time scheduling of flexible customers and PV facilities while considering both the previous uncertainty and uncertainty concerning the operation of PV facilities.

The investigation corresponds with the current policies implemented in the country's distribution sector. For example, it aligns with the flexibility strategy for the SEN concerning the operation scheduling improvement through intra-day updates. Also, to implement the investigation successfully, the separation proposed by the Electrical Portability Bill between electricity distribution, as a business of the electrical grid infrastructure, and electricity trading, as an activity carried out under competition, is crucial.

The proposed framework comprises the following advantages for the DSO: 1) online characterization of customer behavior, thus obtaining mathematical models to estimate the customer response according to consumption preferences and

environmental factors; 2) online scheduling of customers' power responses into the distribution system accounting for uncertainty; 3) optimal determining of inverters' power set points and coordination of their dispatch according to grid requirements and local market conditions; 4) ensuring reliable operation of the distribution system during periods of peak demand or fluctuations of PV production; 5) earning financial benefits; and 6) reducing investment in traditional centralized power reserves by harnessing the combined capacity of the demand-side energy resources.

Finally, the investigation presented in this thesis has been reported in the following relevant publications:

1) L. Marrero, D. Sbárbaro, and L. García-Santander, "Customized Scheduling of Demand Response with Dispatchable Photovoltaic Inverters," *IEEE Transactions on Power Systems* (under review).

2) L. Marrero, D. Sbárbaro, and L. García-Santander, "Online Scheduling of Demand Response for Renewable Integration Considering Uncertainty," *Journal of Modern Power and Clean Energy* (for resubmission).

3) L. Marrero, D. Sbárbaro, and L. García-Santander, "Online Demand Response Characterization Based on Variability in Customer Behavior," *Journal of Modern Power and Clean Energy* (accepted, 2023), doi: 10.35833/MPCE.2023.000516.

4) M. Kippke, L. Marrero, and L. García-Santander, "Advanced Metering Infrastructure for Customers Identification," *Xplore Technology Award 2023 for a Sustainable World*, Bad Pyrmont, Germany, 2023.

5) L. Marrero, L. García-Santander, and D. Sbárbaro, "Caracterización de patrones de consumo y estilos de vida de clientes residenciales," *IX International Congress Biobío Energía*, Concepción, Chile, 2020.

1.4 Thesis Organization

The organization of the thesis is as follows:

Chapter 2 presents the online framework to characterize customers' DR over time. The chapter provides the theoretical foundations and mathematical models and then describes the solution methodology for the online characterization. Two case studies based on real-world and simulated daily profiles test the proposed framework. The chapter concludes with a performance monitoring and a comparison and sensitivity analysis.

Chapter 3 presents the online framework for scheduling customers' DR to support integrating PV generation into the distribution system. The chapter first provides

theoretical foundations and mathematical models. Then, it defines the problem formulation under uncertainty and describes its solution methodology. Finally, a case study with real-world daily profiles of customers and the IEEE-37 node test distribution feeder assesses the proposed approach and analyzes the uncertainty cost.

Chapter 4 presents the online framework designated for the optimal scheduling of the distribution-level energy resources. The chapter introduces the mathematical model for dispatchable inverters of PV facilities and extends the previous problem formulation by considering the uncertainty corresponding to their operation. The proposed framework is assessed in a case study using the real-world daily profiles of customers and the IEEE-37 node test distribution feeder.

Chapter 5 summarizes the contents of the thesis, giving an overview of the most important results. It further provides some directions for future work.

Chapter 2

Online Demand Response Characterization Based on Electricity Consumption Patterns

This chapter proposes an online framework to characterize demand response (DR) over time. The proposed approach facilitates obtaining and updating the daily consumption patterns of customers. The essential concept of response profile class (RPC) is introduced for characterization, complemented by the measure of the variability in customer behavior. The chapter uses for daily profiles a modified version of the incremental clustering by fast search and find of density peaks (CFSFDP) algorithm that considers the multivariate normal kernel density estimator and incremental forms of the Davies-Bouldin (iDB) and Xie-Beni (iXB) validity indices. Case studies conducted using real-world and simulated daily profiles of residential and commercial Chilean end-users demonstrate how the proposed approach can continuously characterize DR. Results prove that the presented framework achieves realistic customer models for effective energy management by estimating the customer response to price signals at the distribution system operator (DSO) level.

2.1 Motivation

The intrinsic socio-demographic characteristics (individual preferences) of customers and real-time externalities (environmental factors) can influence their response [14] in price-based DR programs. Typically, this information is private and unknown. However, understanding how end-users respond conditioned by these influences is essential for estimating their potential for flexibility and designing correct pricing schemes to match the distribution system's operation needs. To that end, a processing and subsequent characterization of daily load profiles are required first. These tasks

represent a notable challenge since electricity data are data streams; thus, online clustering is necessary for handling it.

Although, in general, there are many papers in the literature about online (or stream) clustering foundations and algorithms (see, for instance, the recent surveys [15]-[17]), few studies have been reported directly addressing this topic with application to electricity data, which is the focus of this investigation. The same observation is found in related works [18] and [19].

However, several studies have recently analyzed customers' electricity consumption by exploiting important offline clustering methods. For example, Sun, Konstantelos, and Strbac [20] classify 2613 households under diverse load conditions such as calendar seasons and days. A clustering is applied in [21] to load patterns represented as images, and then periods with similar consumption levels are identified by considering load variation and uncertainty. A bi-level load shape dictionary is developed in [22], where extracted features such as weekly and seasonal patterns and segment entropy characterize customers' energy usage. Consumption dynamics for each end-user are formulated first in [23]; then, from a clustering process, an evaluation of variability in the resulting clusters is performed by an entropy analysis. Haben, Singleton, and Grindrod [24] extract, classify and verify the reliability of the clustering and discover clusters that describe end-users according to their demand and variability. Reference [25] uses an encoding system with a load-shape dictionary in 44 million daily profiles, focusing on segmenting customers' lifestyles. Finally, in [26], a distributed-centralized identification method to extract and characterize typical daily patterns is proposed for industrial customers.

Despite the valuable contributions of these methods, they lack developing an online characterization. On the other hand, De Silva *et al.* [27] present an interim summarization function for the load data streams and an algorithm that incrementally learns and accumulates characterized patterns in six smart meters (SMs). Reference [28] implements the division of the load data streams in time windows, where objects in each window are assigned to different clusters (or concepts) whose structural change over time is analyzed, although they are not updated recursively. Similarly, an approach is presented in [19] that, by extracting interpretable features from consumer data, performs change detection and improves forecasting on aggregated time series within clusters. An incremental algorithm is devised in [18] to detect pattern drifts through load pattern extraction, intergradation, and modification, but it works with only one customer simultaneously. Finally, Le Ray and Pinson [29] introduce an online adaptive clustering for load profiling. However, these studies mainly focused on analyzing changes in consumption patterns rather than characterizing customer

behavior. The main advantage of this characterization for the DSO is obtaining online mathematical models to estimate the customer response according to consumption preferences and environmental factors. The DSO can then procure electrical energy with more certainty from the electricity pool.

This chapter presents an online framework for characterizing the customer response from the DSO perspective. The continuous processing of daily profiles makes it possible to know the customers' RPCs and compute their variability. The methodology comprises a modified version of the incremental CFSFDP algorithm of Zhao *et al.* [30], based on the seminal work of Rodriguez and Laio [31]. Previous experiences using the CFSFDP algorithm for load profiling in related studies [23] and [26] support this selection. The incremental formulation adapts this original offline algorithm to work in an online setup. The presented version considers the multivariate normal kernel density estimator, which is robust to the number of objects in clusters during online processing, and the iDB and iXB validity indices [32] to provide information about the algorithmic performance. The DSO can obtain the expected consumptions of customers from their corresponding RPCs.

The customer response to the price signal has also been studied recently in the literature in the context of DR pricing. For example, in [33], strategies for setting real-time prices are developed by implicitly learning consumer's price elasticity online, although only own-price elasticity is in load changes. In [34], a model from the aggregator perspective runs a pricing program with distribution system constraints and learns the price sensitivities of customers. Finally, Tao *et al.* [35] adopt a non-intrusive load monitoring-based pricing approach that estimates the DR potential of thermostatically controlled loads and then models the price responsiveness of customers. While how to generate time-varying price signals is beyond the scope of this chapter, the proposed approach provides the DSO with two powerful instruments to achieve this: 1) the updated customer models for the estimation of responses to different price signals; 2) the modeling of the underlying probability distribution of random deviations in demand from expected values, inherent to the customer' stochastic behavior. According to [9], demand uncertainty generally has a normal distribution; therefore, the approach is suitable for finding information about the statistical moments.

The main contributions of the chapter are summarized below:

1) An innovative framework for online DR characterization is proposed. The chapter presents the modified incremental CFSFDP algorithm, defined to work in a Hilbert space. Online clustering introduces the multivariate normal kernel density estimator

for robustness to the number of objects in clusters and the monitoring of algorithmic performance through the iDB and iXB validity indices.

2) Application of the proposed approach allows the DSO to perform two essential activities: updating the RPC and variability when customer response materializes (at the end of the day) and estimating the customer behavior to price signals based on a known RPC (within the next day).

3) The framework is tested with real-world and simulated daily profiles. Results show the online process for obtaining RPCs of residential and commercial Chilean end-users. The chapter also provides a comparison analysis with the online algorithm in [29] and a sensitivity analysis employing different combinations for the number of representatives and shrink factor.

The organization of the chapter is as follows. Section 2.2 provides the theoretical foundations and mathematical models. Section 2.3 describes the solution methodology. Section 2.4 presents two case studies with real-world and simulated daily profiles to assess the proposed framework. The performance monitoring and the comparison and sensitivity analysis are discussed in Section 2.5. Finally, Section 2.6 concludes the chapter.

2.2   Problem Framework

The DSO needs to ensure the reliability of the distribution grid, which may include small distributed solar generation units and generally has tight capacity constraints. By appropriately choosing dynamic price signals to be broadcasted to consumers enrolled in a price-based DR program, the DSO can reduce distribution system costs and increase reliability, for example, by shifting flexible consumption to periods with high stochastic production [2].

This chapter considers the price-setting DSO that aims at managing the demand flexibility and pursues estimating the consumers' behavior to price signals. To this end, the DSO performs the daily processing of load profiles through the modified incremental CFSFDP algorithm to update the RPCs and variability of customers. Using the corresponding RPCs, the DSO can generate the expected consumption profiles in response to control price signals. Therefore, it can decide with high certainty how much electricity to trade, for example, in the balancing market within the day. Figure 2.1 illustrates the problem framework.
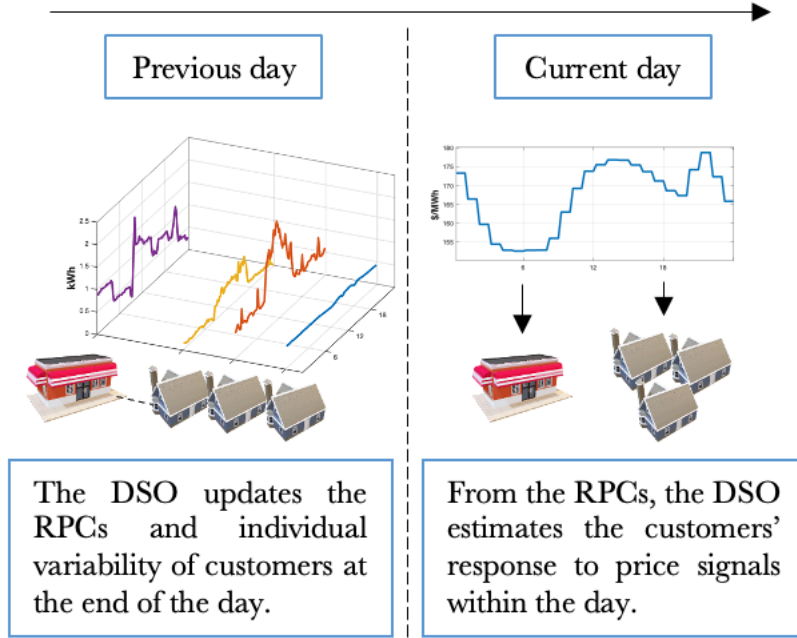
Figure 2.1. Framework for the online DR characterization.

2.2.1 Flexible Customer Expected Response

Let $\bar{s}_{l_t} = \bar{p}_{l_t} + j\bar{q}_{l_t}$ designate an expected complex power value to be consumed at time point t by a customer l under a contract; therefore, $\bar{p}_{l_t} = \Re\left(\bar{s}_{l_t}\right)$ is the active component and $\bar{q}_{l_t} = \Im\left(\bar{s}_{l_t}\right)$ the reactive one. Furthermore, $\bar{s}_{l_t} \in \mathcal{S}_{l_t}$, where $\mathcal{S}_{l_t}$ is the consumption region that contains the customer's active and reactive power values in t, positioned in the complex plane. It is possible to obtain a bounded and convex approximation $\mathcal{S}'_{l_t}$ of this region given practical bounds, both for active and reactive power. Although the estimation of $\bar{q}_{l_t}$ is fundamental for the analysis at the distribution system level, this chapter focuses specifically on the active power responses of end-users. The following linear model for setting a flexible active power profile is defined [2]:

$$p_{l_t}^{\min} \leq \bar{p}_{l_t} \leq p_{l_t}^{\max} : \beta_{l_t}^{\min}, \beta_{l_t}^{\max}, \quad l \in \mathbf{L}, t \in \mathbf{T}, \tag{2.1}$$

$$-r_{l_t}^{d}\Delta t \leq \bar{p}_{l_t} - \bar{p}_{l_{t-1}} \leq r_{l_t}^{u}\Delta t : \psi_{l_t}^{d}, \psi_{l_t}^{u}, \quad l \in \mathbf{L}, t = 2,\dots,T, \tag{2.2}$$

$$\sum_{t=1}^{T} \bar{p}_{l_t}\Delta t \geq e_l : \varepsilon_l, \quad l \in \mathbf{L}, \tag{2.3}$$

Equation (2.1) provides the expected response $\bar{p}_{l_t}$ between a minimum $p_{l_t}^{\min}$ and maximum $p_{l_t}^{\max}$ bound for customer l at time point t, where $\bar{p}_{l_t}$ is an element of vector $\bar{\mathbf{p}}_l$ that represents the expected profile of the customer, that is, $\bar{\mathbf{p}}_l = [\bar{p}_{l_1} \ \cdots \ \bar{p}_{l_T}]^\top$. Likewise, $\mathbf{L} = \{1, \dots, L\}$ and $\mathbf{T} = \{1, \dots, T\}$ are respectively the sets of customers and time points within the day. Also, $\bar{p}_{l_t}$ can increase or decrease depending on the market price due to the combined use of shifting and shedding loads. Equation (2.2) forces ramp limits on the decrease and increase of active power in two successive time points, with $r_{l_t}^{\mathrm{d}}$ and $r_{l_t}^{\mathrm{u}}$ as the maximum ramp-down rate and ramp-up rate, respectively. Finally, a minimum daily energy e_l is specified by (2.3) to account for basic activities, where Δt is the interval between two consecutive time points. Variables arranged after the colon are dual.

An important observation to consider actual consumption features of consumers is that each region $\mathcal{S}'_{l_t}$ is time-varying since the bounds vary over time based on their preferences and environmental factors. The study exploits this and develops an online processing and subsequent characterization of daily load profiles. From the corresponding outcome, it is attainable to differentiate the behaviors of consumers through RPCs, where each RPC represents for a customer a portion (of similar daily profiles or vectors) of the polytope that entirely contains its load scenarios in the vector space. Therefore, based on these classes or portions, a more refined estimation of the consumption activity is feasible.

From a set of daily profiles associated with a RPC, each pair of parameters $p_{l_t}^{\min}$ and $p_{l_t}^{\max}$ of the model can be obtained as the corresponding extreme values, providing the convex (inner) approximation of the active power.

Regarding the values of the maximum ramp rates in (2.2), because they are related to the speed at which the consumer can decrease or increase its demand, they differ for each time point of the day and between RPCs. The strategy for its online determination is to consider the load changes from time point $t - 1$ to the next t, $t = 2, \dots, T$, within the set of daily profiles of the RPC. Then, the following expressions result:

$$r_{l_t}^{\mathrm{d}}\Delta t = \max_{\{p_{l_{t-1}}-p_{l_t}\}} \left(p_{l_{t-1}} - p_{l_t}\right), \quad l \in \mathbf{L}, t = 2, \dots, T, \quad (2.4)$$

$$r_{l_t}^{\mathrm{u}}\Delta t = \max_{\{p_{l_t}-p_{l_{t-1}}\}} \left(p_{l_t} - p_{l_{t-1}}\right), \quad l \in \mathbf{L}, t = 2, \dots, T, \quad (2.5)$$

where (2.4) and (2.5) indicate a decrease and an increase in demand concerning the previous time point, respectively.

Lastly, the minimum daily energy can be the lowest total consumption among all the profiles within the RPC.

2.3 Solution Methodology

Daily processing of load profiles favors the appropriate characterization of customer behavior. To make this analysis scalable and because consumers can be associated based on the similarity of their consumption patterns, this work presents an incremental clustering, whose description is the first focus of the section. The rest considers the performance monitoring and the estimation of the variability of the customer response.

Following Figure 2.1, **Algorithm 1** outlines the online workflow that runs at the end of each day when customer response materializes.

2.3.1 Modified Incremental CFSFDP Algorithm

Let $\mathbf{P}_0 = \{\mathbf{p}_l^0\}$, with each vector $\mathbf{p}_l^0 = [p_{l\,1}^0 \; \cdots \; p_{l\,\tau T}^0]^\top$, denote an initial set of power profiles collected during τ days from L customers equipped with SMs. Let the initial power profile of each customer recast into τ daily load profiles. By gathering all these new profiles, the initial set is reformulated then with a total of $N_0 = \tau L$ vectors of T-tuples, that is, $\mathbf{P}_0 = \{\mathbf{p}_i^0, i = 1, \ldots, N_0\}$, with each vector $\mathbf{p}_i^0 = [p_{i\,1}^0 \; \cdots \; p_{i\,T}^0]^\top$. Furthermore, a set $\mathbf{P} = \{\mathbf{p}_l\}$ of load profiles is processed daily after this historical collection.

Any structure in the vector space depends on the similarity metric and the clustering criterion, which expresses how to use the metric. In the investigation, the metric $d_2 \colon \mathbb{R}^T \times \mathbb{R}^T \to \mathbb{R}$, defined in (2.6) in terms of the ℓ_2 norm for any pair of vectors $\mathbf{p}_i^0$ and $\mathbf{p}_j^0$, is employed. Hence, $\mathbb{R}^T$ is a normed linear vector space, particularly a Hilbert space, due to the induced norm [36].

$$d_2\left(\mathbf{p}_i^0, \mathbf{p}_j^0\right) = \left\|\mathbf{p}_i^0 - \mathbf{p}_j^0\right\|_2, \quad \forall\, \mathbf{p}_i^0, \mathbf{p}_j^0 \in \mathbf{P}_0, \qquad (2.6)$$

The implementation of the modified incremental CFSFDP algorithm involves the following stages:

1) Application of the CFSFDP Algorithm: The CFSFDP algorithm uses the initial set $\mathbf{P}_0$, where the clustering criterion relies on the computation of two magnitudes for each object $\mathbf{p}_i^0$: the local density ρ_i and the minimum distance δ_i concerning the vectors of higher density:

$$\rho_i = \frac{1}{N_0 h_1 \cdots h_T} \sum_{j=1}^{N_0} \left[\prod_{t=1}^{T} \mathscr{k} \left(\frac{p_{i\,t}^0 - p_{j\,t}^0}{h_t} \right) \right], \quad i = 1, \dots, N_0, \quad (2.7)$$

$$\delta_i = \min_{\mathbf{p}_j^0 \in \mathbf{P}_0} d_2\left(\mathbf{p}_i^0, \mathbf{p}_j^0\right) \qquad (2.8)$$

$$\text{s.t. } \rho_j \geq \rho_i, \quad i, j = 1, \dots, N_0, \qquad (2.9)$$

where, unlike [30] and [31] that use the cutoff distance of high sensitivity to a small number of objects, (2.7) denotes the general form of a multivariate product kernel estimator at point $\mathbf{p}_i^0$ of the probability density function, in which the same (univariate) kernel function $\mathscr{k}(\cdot)$ is used each time point t with a different smoothing parameter h_t. Without loss of generality, let the normal kernel be selected. Then, according to Scott's rule [37] in $\mathbb{R}^T$, each smoothing parameter can be approximated as $h_t = \sigma_t N_0^{-1/T+4}$, with σ_t as the standard deviation for time point t.

For the object $\mathbf{p}_i^0$ with the highest density, the distance is:

$$\delta_i = \max_{\mathbf{p}_j^0 \in \mathbf{P}_0} d_2\left(\mathbf{p}_i^0, \mathbf{p}_j^0\right) \qquad (2.10)$$

Each cluster center $\mathbf{c}_k^0$, $k = 1, \dots, K$, describes a dominant consumption pattern and is expected to be surrounded by a neighborhood with lower local density and at a relatively large distance from any object with a higher local density. To identify the centers, both the plot of δ_i as a function of ρ_i for each object (the so-called decision graph) and the plot of quantities $\gamma_1, \dots, \gamma_{N_0}$ in decreasing order, each $\gamma_i = \rho_i \delta_i$, can be used [31]. After detecting the centers, each remaining object is assigned to the same cluster as its nearest neighbor of higher density. Each cluster is denoted as $\mathbf{G}_k^0 = \left[\mathbf{c}_k^0 \; \mathbf{p}_{k,2}^0 \; \cdots \; \mathbf{p}_{k,N_k^0}^0\right]$, where N_k^0 is the number of daily profiles.

2) Determination of the Representatives: With this approach, each cluster is represented by a fixed number of representative points generated by selecting first well-scattered points and then shrinking them toward the center by a specified fraction. The approach helps to identify clusters with non-spherical shapes and wide variances in size [38].

Let N_r be the number of representative points of clusters and let $\mathbf{R}_k^0 = \{\mathbf{r}_{k,u}^0, u = 1, \dots, N_r\}$ be the set of these points to select for any cluster k. Their determination follows a sequential order. A selected vector $\mathbf{p}_{k,i}^0$ becomes a representative vector $\mathbf{r}_{k,u}^0$. The first representative is:

$$\mathbf{r}_{k,1}^0 = \arg\max_{\mathbf{p}_{k,i}^0 \in \mathbf{G}_k^0} d_2(\mathbf{p}_{k,i}^0, \mathbf{c}_k^0) \tag{2.11}$$

while the rest of them are selected one by one as follows:

$$\mathbf{r}_{k,u}^0 = \arg\max_{\mathbf{p}_{k,i}^0 \in \mathbf{G}_k^0 \backslash \mathbf{R}_k^0} \left[\min_{\mathbf{r}_{k,v}^0 \in \mathbf{R}_k^0} d_2\left(\mathbf{p}_{k,i}^0, \mathbf{r}_{k,v}^0\right) \right] \tag{2.12}$$

$$\text{s.t. } d_2\left(\mathbf{p}_{k,i}^0, \mathbf{r}_{k,v}^0\right) \geq e^{-s(\mathbf{G}_k^0)}, \quad k = 1, \dots, K, u = 2, \dots, N_r, \tag{2.13}$$

where $\mathbf{r}_{k,v}^0$ is a representative vector in $\mathbf{R}_k^0$, for instance, to select the second representative, $\mathbf{r}_{k,v}^0$ is $\mathbf{r}_{k,1}^0$, to select the third representative, $\mathbf{r}_{k,v}^0$ could be $\mathbf{r}_{k,1}^0$ or $\mathbf{r}_{k,2}^0$, and so on; and $s(\mathbf{G}_k^0)$ is the structure distance of $\mathbf{G}_k^0$, expressed in (2.14) based on the mean $\mu_{k,t}$ and standard deviation $\sigma_{k,t}$ of time point t [30].

$$s(\mathbf{G}_k^0) = \sqrt{\sum_{t=1}^{T} \left(\frac{\sigma_{k,t}}{\mu_{k,t}}\right)^2 \frac{\sum_{t=1}^{T} \mu_{k,t}}{T}}, \quad k = 1, \dots, K, \tag{2.14}$$

The shrinking process of the representative points depends on the (user-defined) shrink factor $\alpha \in [0,1]$:

$$\mathbf{r}_{k,u}^0 = \mathbf{r}_{k,u}^0 + \alpha\big(\mathbf{c}_k^0 - \mathbf{r}_{k,u}^0\big), \quad k = 1, \dots, K, u = 1, \dots, N_r, \quad (2.15)$$

Shrinking the scattered points toward the center undoes surface abnormalities and mitigates the effect of outliers since these are typically further away from the cluster center [38].

3) Assignment of New Load Profiles: With a new daily set $\mathbf{P}$, each object $\mathbf{p}_l$ is assigned to the cluster $\mathbf{G}_k^0$ with the nearest representative. To that end, the local density ρ_l is set initially to zero, and the minimum distance δ_l is as follows:

$$\delta_l = \min_{\mathbf{r}_{k,u}^0 \in \mathbf{R}_k^0} d_2\big(\mathbf{p}_l, \mathbf{r}_{k,u}^0\big) \qquad (2.16)$$

The assignment causes a change in the cluster structure, which requires updating the representative points. Assuming $\mathbf{r}_{k,v}^0$ as the nearest representative to the assigned vector $\mathbf{p}_{k,l}$, the solution of the problem below is found:

$$\arg \max_{\mathbf{p}_{k,l}, \mathbf{r}_{k,v}^0 \in \mathbf{A}} \left(\min_{\mathbf{r}_{k,u}^0 \in \mathbf{R}_k^0 \backslash \mathbf{r}_{k,v}^0} d_2\big[\mathbf{A}(1:T,c), \mathbf{r}_{k,u}^0\big] \right) \qquad (2.17)$$

$$\text{s.t. } d_2\big[\mathbf{A}(1:T,c), \mathbf{r}_{k,u}^0\big] \geq e^{-s(\mathbf{G}_k^0)}, \quad k = 1, \dots, K, l \in \mathbf{L}, c = 1,2, \quad (2.18)$$

where $\mathbf{A} = \big[\mathbf{p}_{k,l} \ \mathbf{r}_{k,v}^0\big]$ is an auxiliary matrix, with the column c.

From this result, if $\mathbf{p}_{k,l}$ produces the maximum value, then $\mathbf{p}_{k,l}$ replaces $\mathbf{r}_{k,v}^0$ as the new representative [30]. Each cluster is denoted as $\mathbf{G}_k = \big[\mathbf{c}_k^0 \ \mathbf{p}_{k,2} \ \dots \ \mathbf{p}_{k,N_k}\big]$, where N_k is the number of daily profiles.

4) Splitting Procedure: In this stage, the algorithm looks for clusters with more than one dominant pattern. Specifically, in each new cluster $\mathbf{G}_k$, parameters ρ_i and δ_i, and the quantity γ_i, are first obtained for each object $\mathbf{p}_{k,i}$. Based on these product values, new cluster centers can be identified, and the remaining objects assigned as in the CFSFDP algorithm [30]. Since only very few clusters arise in practice, an empirical criterion by computing the mean of the ten highest quantities can deliver new centers if these objects have a value higher than this.

5) Merging Procedure: This stage happens if new clusters arise, looking for those containing a similar dominant pattern. Specifically, the connected graph is constructed to find components connected between clusters. The minimum distance between any two clusters $\mathbf{G}_k$ and $\mathbf{G}_m$, $k, m = 1, \dots, K$, is computed as:

$$D(\mathbf{G}_k, \mathbf{G}_m) = \min_{\mathbf{r}_{k,u} \in \mathbf{R}_k, \mathbf{r}_{m,v} \in \mathbf{R}_m} d_2(\mathbf{r}_{k,u}, \mathbf{r}_{m,v}) \qquad (2.19)$$

If $D(\mathbf{G}_k, \mathbf{G}_m) \leq s(\mathbf{G}_k)$ and $D(\mathbf{G}_k, \mathbf{G}_m) \leq s(\mathbf{G}_m)$, an edge is added between them. After adding all the edges, the graph with multiple components results, and the clusters with the same component can be merged [30].

2.3.2 Incremental Indices for Performance Monitoring

Lately, some authors [32], [39], [40] have extended several of the offline cluster validity indices to deal with the clustering over data streams. This study implements the iDB and iXB indices to monitor the performance of the online algorithm.

For indices' computation, the compactness term is essential. The study considers the incremental formulation proposed in [32], which represents a counterpart hard for the calculation of the compactness as considered in [39] (in the context of fuzzy clustering). This formulation is applied to clusters that do not undergo splitting or merging, and for calculating both indices, the corresponding equations in [32] are employed (see Appendix A). Instead, the work applies the offline forms of both indices to clusters that arise due to a split or merger. However, the compactness formulation in this study is more straightforward than the one described in [32] due to the following differences: 1) the calculation is executed after the assignment of objects in $\mathbf{P}$ to clusters and not after the assignment of each, 2) the use of a cluster center $\mathbf{c}_k$ instead of the centroid, which changes with the added objects. In both indices, smaller values mean better solutions, whereas sudden changes indicate changes in the cohesion and separation of clusters produced by the online algorithm [17].

2.3.3 Variability of the Customer Response

To measure the uncertainty about the daily RPC for each customer, the entropy value [36] is used in the study:

$$H_l = -\mathbb{E}[\log \mathbb{P}(\mathbf{c}_k)], \quad l \in \mathbf{L}, k = 1, \dots, K, \qquad (2.20)$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator and $\mathbb{P}(\cdot)$ is the probability of the cluster center followed up to the current day.

Equation (2.20) is solved after each daily observation of RPCs and gives the amount of information inherent to each, that is, the variability in customer behavior.

Algorithm 1 Online DR Characterization

Input: Set $\mathbf{P}$, $\mathbf{G}_k^0$, $\mathbf{R}_k^0$, $k = 1, \dots, K$

1: **for** $t \leftarrow T$ **do**
2:     Assign each object $\mathbf{p}_l$ to the corresponding cluster by (2.16)
        and solve $(2.17) - (2.18)$ to update the representative
3:     **for** each cluster $\mathbf{G}_k$, **do**
4:         Compute the local density by (2.7) and the minimum
            distance by $(2.8) - (2.10)$ for each object, and select
            new cluster centers through their products if they exist
5:         Assign each remaining object to the corresponding cluster
            and select the representatives by $(2.11) - (2.15)$
6:     **end for**
7:     **if** new clusters arise **then**
8:         Compute the minimum distance between all pairs of
            clusters by (2.19) and merge accordingly
9:         In the merged clusters, compute the local density by (2.7)
            and the minimum distance by $(2.8) - (2.10)$ for each
            object, select as center the object with the highest product
            value and the representatives by $(2.11) - (2.15)$
10:    **end if**
11:    Compute the iDB and iXB validity indices
12:    Update the RPCs and variability of customers by (2.20)
13: **end for**

2.4 Case Studies

This section presents two case studies to demonstrate the benefits of the approach. The first case study involves residential and commercial electricity data recorded over six weeks (February 1 to March 13, 2020), with 15-minute intervals. The Hilbert space is then $\mathbb{R}^{96}$. The number of customers is 925, charged with a regulated tariff; however, the study assumes they practice an optimizing behavior, using electricity in known off-

peak periods. Therefore, **Algorithm 1** is applied continuously when each day ends, simulating the succession of days during this period. As a result, the DSO can update the RPCs and variability of customers at the end of the day. The second case study considers numerical simulations for an additional week, where the DSO broadcasts price signals daily. Based on each daily update of RPCs from **Algorithm 1**, this case study allows obtaining expected load profiles of customers for the next day.

2.4.1 Case study Using Real-World Data Set

Incompleteness in electricity data is a common trend. Then, cleaning is executed by identifying and discarding daily profiles with missing and inconsistent values. The initial set $\mathbf{P_0}$ considers the first week of measurement data. Thus, each set $\mathbf{P}$ of profiles corresponding to the rest of the days is processed incrementally. Likewise, the normalization of each daily profile concerning its maximum value is implemented to facilitate the clustering process.

The application of the CFSFDP algorithm to $\mathbf{P_0}$ allows the identification of four initial clusters through the decision graph and the plot of quantities $\gamma_1, \ldots, \gamma_{N_0}$ in decreasing order. Figure 2.2 depicts this result and highlights the selected centers with colored and bigger dots. One of them is remarkably different, while the remaining three differ from the rest mainly due to the distance parameter, which means that these points do not have a neighborhood as high as the first.
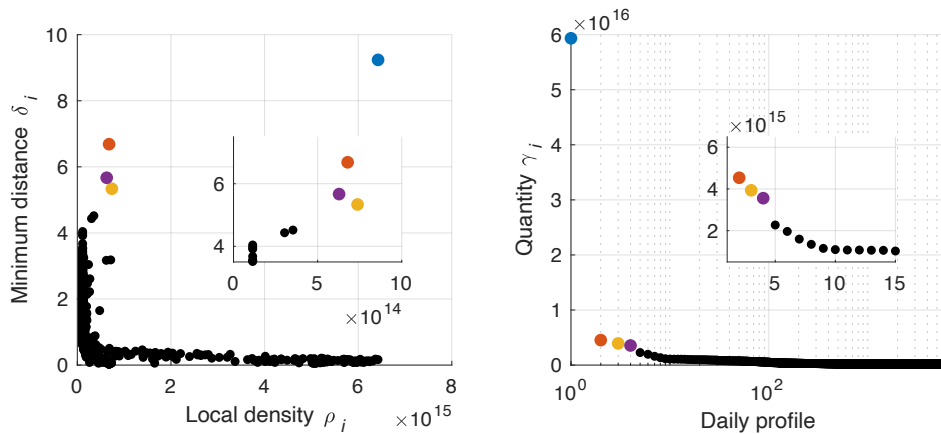
Figure 2.2. Plots for identification of the initial clusters, including a zoom to better distinguish the last three: the decision graph and quantities $\gamma_1, \ldots, \gamma_{N_0}$ in decreasing order (with abscissa axis in logarithmic scale).

Representative points in clusters are the basis for assigning new profiles. The incremental clustering uses the following parameter values: $N_r = 8$ and $\alpha = 0.4$. To better observe the shrinking of representatives, Figure 2.3 shows the process for the points that belong to the initial cluster 3, with the lowest cardinality, into a two-dimensional space using the principal component analysis (PCA) [36]. Representative points and the cluster center are highlighted in bigger black dots and a blue diamond, respectively. In $\mathbb{R}^{96}$, this comprises the translation of the points around the center.
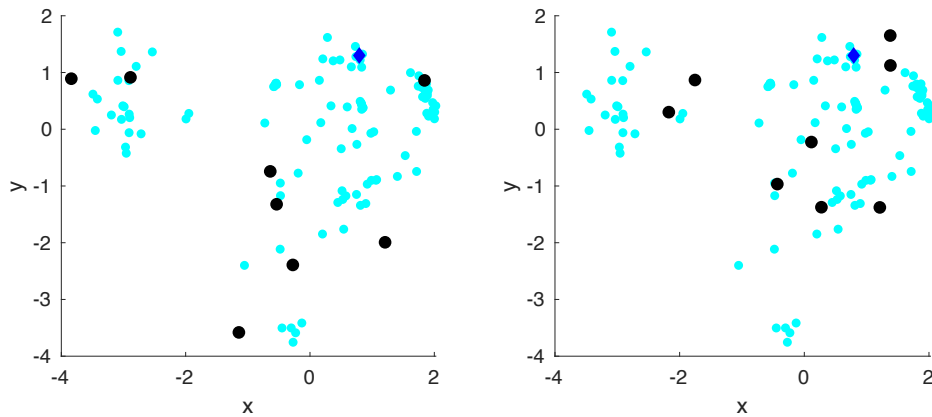
Figure 2.3. Scatter data of cluster 3 for PCA before and after the shrinking process, with representatives in bigger black dots and the cluster center in a blue diamond.

The online algorithm produces five final clusters considering the real-world data set. Figure 2.4 illustrates its evolution starting from the initial four. Following the trajectories, it is possible to identify the origin for each cluster, that is, the one from which they arise and the corresponding date. In the same way, for those clusters that fade, it is noticeable which one they merge with and the corresponding date. Cluster 1 is the only one that maintains from the beginning. Also, several arise and fade in the last days of the period, which is attributed to a change in the consumption behavior of most customers since February is a vacation month and March is a month of more work activity and back to school. Finally, some isolated objects (with small density and great distance) are classified as outliers and discarded during online processing.

Since customers generally have well-defined behaviors (at least for a specific period) and these behaviors repeat between them, continuous splitting of clusters is uncommon. Also, a splitting rarely generates more than two clusters. The same happens with merging. Figure 2.4 confirms these observations, with some exceptions in the last days where splits and mergers increase. On the other hand, since a cluster

might contain at most two dominant patterns, the empirical criterion presented in the splitting procedure to split it directly (without analysis as in Figure 2.2) is simple but effective (this includes the isolated objects treated as outliers).
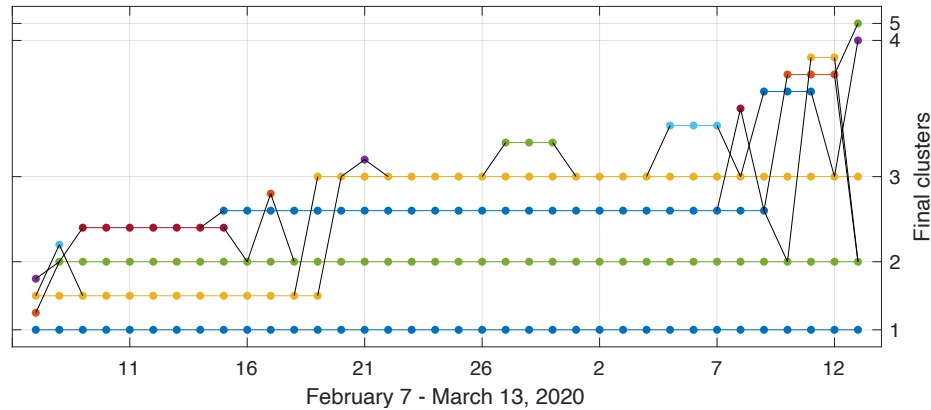
Figure 2.4. Evolution of clusters for the measurement period.

Figure 2.5 depicts the response profiles for the five clusters, and Table 2.1 gives their cardinality and daily average consumption. The most increased percentage of cluster 2 (over 70 %) concludes that this pattern is present in most residential end-users; however, it also includes commercial establishments and small businesses. Cluster 1 presents the highest consumption and more stable behavior. Cluster 3 shows a slightly lower consumption in the morning. Daily profiles with irregular behavior are much more noticeable in cluster 4, and cluster 5 has a typical residential pattern of low consumption.

Table 2.1. Cardinality and Daily Average Consumption of the Final Clusters

Final cluster	Cardinality	Average consumption (kWh)
1	4153	50.5
2	26684	30.3
3	2643	37.4
4	3629	27.8
5	349	19.9

The appropriate characterization of customers' behavior is essential in a control-by-price strategy. Table 2.2 summarizes the RPCs of customers and the combinations of these daily patterns, which is the main benefit of the present framework. Most consumers use four RPCs within the measurement period.
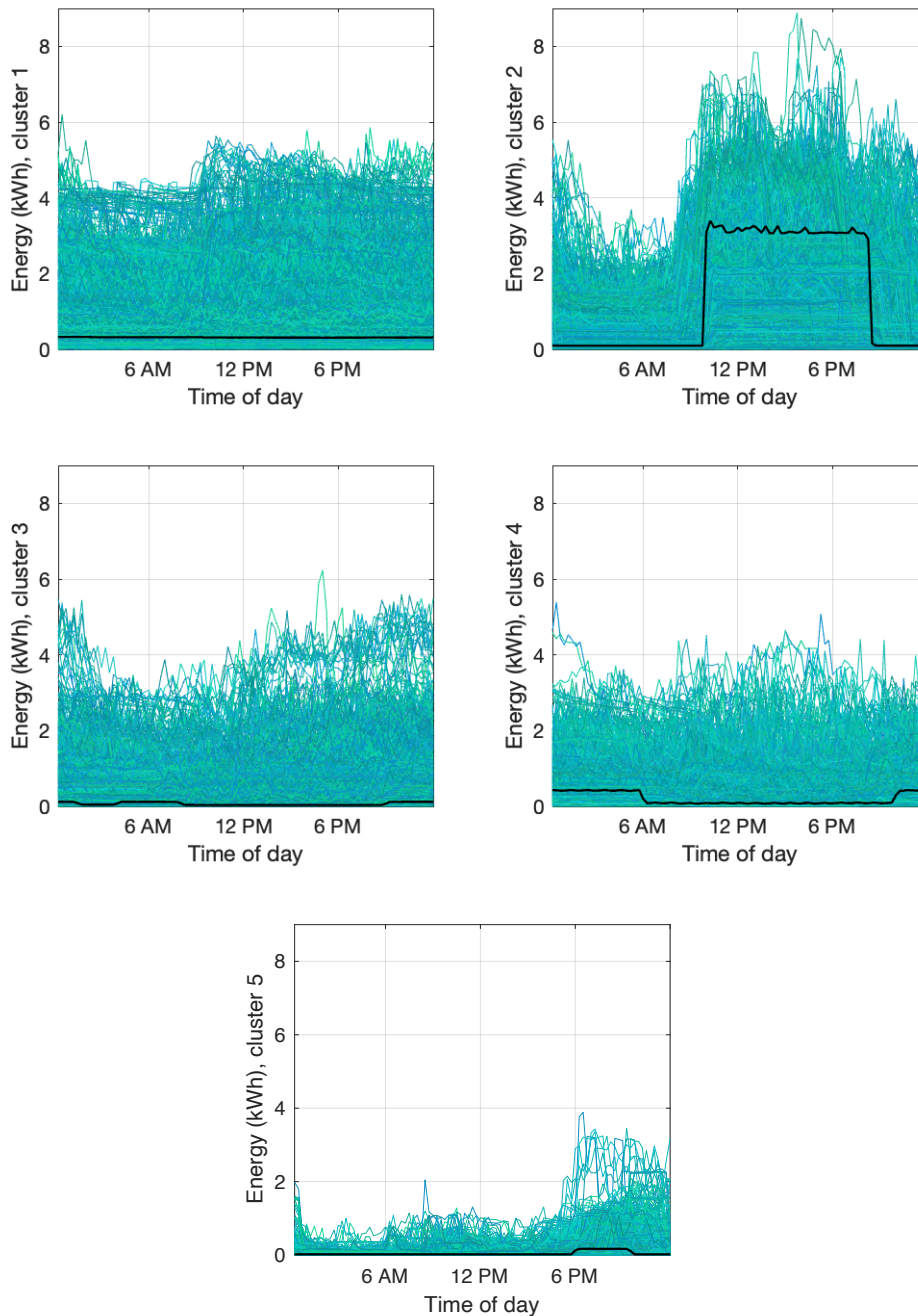
Figure 2.5. Daily consumption profiles of the final clusters, with cluster centers in black.

To complement this, the variability between the patterns in the combination is fundamental. For example, customers 99 and 630 (these are internal identifications for privacy-preserving) have equally the most common combination: 1-2-3-4; however, the probabilities on energy usage are very different, which are 0.047-0.905-0.024-0.024 and 0.286-0.238-0.286-0.19, respectively, and their corresponding entropy values are $H_{99} = 0.1796$ and $H_{630} = 0.5965$ (with the general mean of 0.2436). The lower value is because of the predominance of the second pattern and the very low probability of the rest, which implies less uncertainty about the daily RPC for customer 99. Figure 2.6 confirms this information based on the daily consumption profiles of both consumers, where a more uncertain daily behavior characterizes customer 630.

Table 2.2. RPCs of Customers and their Combinations

RPC	1	2	3	4	5
Number of customers	130	215	264	294	22
Combination of RPCs (number of customers of the combination)	2 (122)	1-2 (85)	1-2-3 (58)	1-2-4-5 (14)	1-2-3-4-5 (22)
	4 (4)	2-4 (100)	1-2-4 (69)	1-2-3-4 (248)	
	1 (4)	1-4 (2)	2-4-5 (32)	2-3-4-5 (29)	
		2-3 (8)	2-3-4 (91)	1-2-3-5 (3)	
		1-3 (6)	1-3-4 (6)		
		2-5 (12)	2-3-5 (4)		
		3-4 (2)	1-2-5 (4)		

2.4.2 Case Study Using Simulated Data Set

This case explores the effect of relaxing the regulatory condition to allow the price to fall in off-peak periods and to increase in peak periods. For simplicity, the DSO broadcasts a single price signal to all consumers each day of the specified week (from March 14 to 20, 2020); however, customized price signals can be designed according to customers' behavior and broadcasted, for example, each hour of the day to exploit newly available information of the system states. Figure 2.7 depicts these price signals generated from the hourly demand profiles of the power system in that week [41].
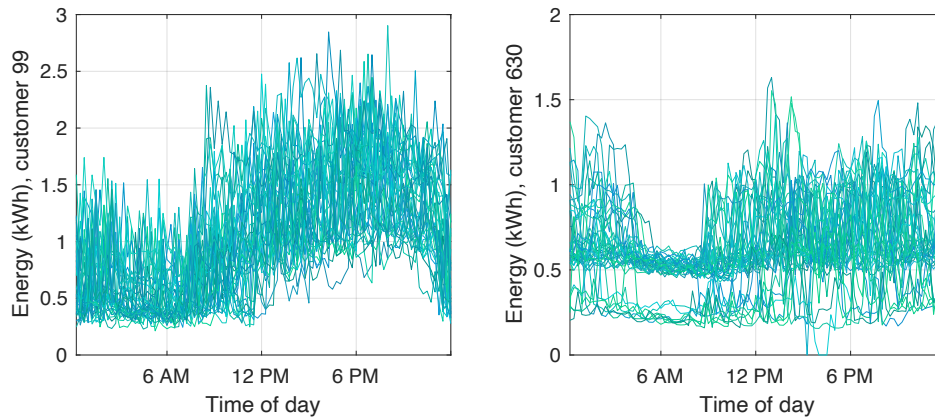
Figure 2.6. Daily consumption profiles of customers 99 and 630 for the measurement period.
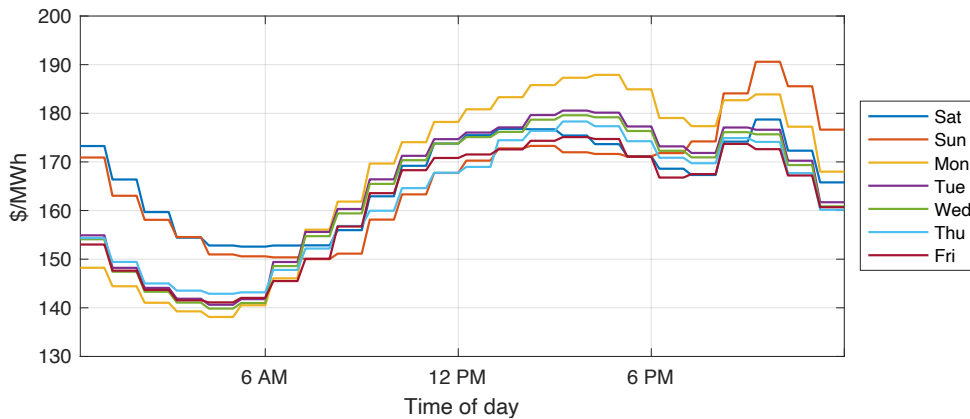
Figure 2.7. Price signals generated according to the Chilean power system demand from March 14 to 20, 2020.

The execution of numerical simulations considers the following ideas: 1) a realistic situation in which consumers use each day the RPC with the highest probability for that day; 2) to estimate active power responses to the price signal, customers practice a daily cost minimization; hence, the chapter employs the linear programming problem in (2.21) [2], where λ_{l_t} represents the electricity price at time point t; 3) to account for the presence of stochasticity in the responses, each random deviation from the expected active power value at time point t follows a normal distribution with mean zero and variance of 0.01 kWh^2.

$$\min_{\bar{\mathbf{p}}_l} \left[\sum_{t=1}^{T} \lambda_{l_t} \bar{p}_{l_t} \Delta t \right] \tag{2.21}$$

$$\text{s.t.} \ (2.1) - (2.3)$$

For example, considering the same two customers and March 14, the RPC with the highest probability for this day results in RPC 2 for customer 99 and RPC 3 for customer 630. Figure 2.8 represents their corresponding mathematical models according to the RPCs, from which the optimizing behavior can be found. Each model involves the minimum and maximum bounds, which in general follow the shape of clusters 2 and 3 in Figure 2.5, respectively, and the maximum ramp values given by products $r_{l_t}^{\mathrm{d}} \Delta t$ and $r_{l_t}^{\mathrm{u}} \Delta t$. Both consumers present time points with a high predominance of one of the ramp rates, which implies a higher flexibility of that ramp rate. For customer 630, lower values around 6 AM indicate less flexibility for this time. Finally, the minimum daily energy is **73.73 kWh** for customer 99 and **34.45 kWh** for customer 630.

After the simulation period, the final clusters increase to six. Figure 2.9 illustrates their response profiles, and Table 2.3 gives their cardinality and daily average consumption. The main difference concerning the case with the real-world data set is the new arising cluster 5, related in shape to the previous cluster 2. Most of the new daily profiles under the effect of the price signals are added to this cluster, which is also indicated in Table 2.3. Considering the shape of the control signals with the lower prices in the early morning and the corresponding constraints on bounds and maximum ramp rates during these hours, most of the new daily profiles present lower consumption to reduce the final cost. In addition, the algorithm assigns the lower consumption profiles of cluster 2 to this new cluster. Based on these factors, cluster 5 shows the lowest energy value, whereas for cluster 2, the cardinality decreases notably, and its average consumption increases. Finally, the same cluster centers hold from the case using the real-world data set.

Table 2.4 summarizes the RPCs and their combinations including the simulation period. Most consumers continue to use four RPCs, but the most common combination is now 1-2-3-4-5.
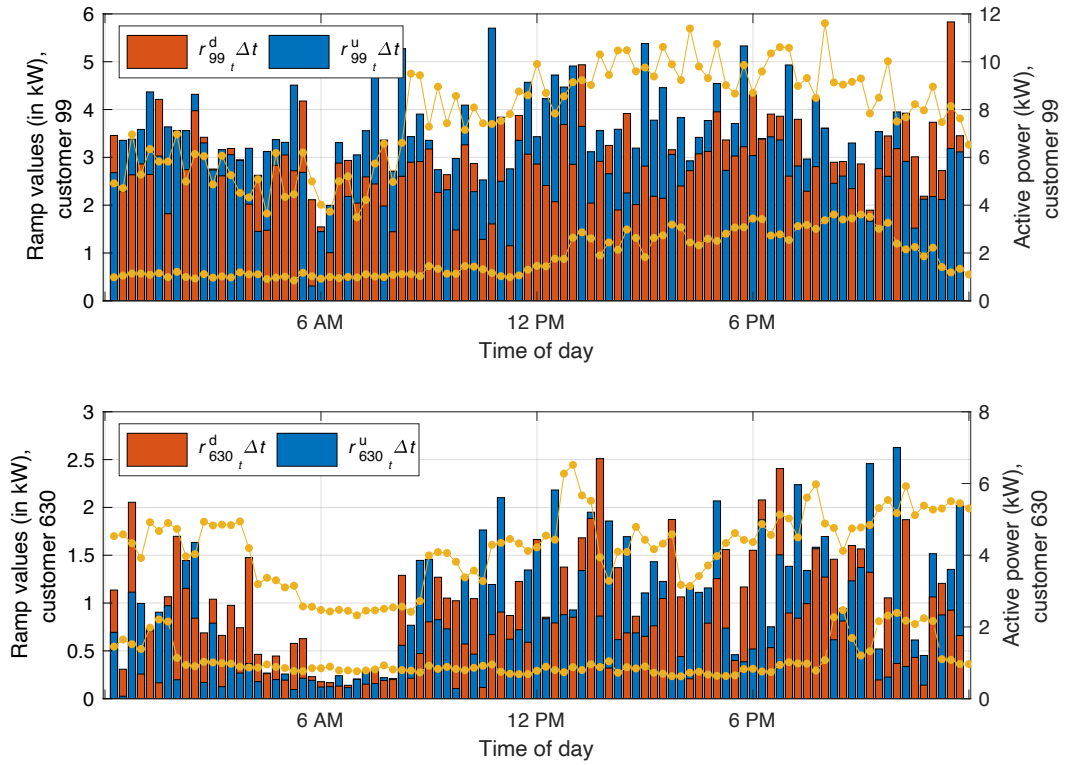
Figure 2.8. Minimum and maximum bounds (in green, right scale) and ramp values $r_{l_t}^{\mathrm{d}}\varDelta t$ and $r_{l_t}^{\mathrm{u}}\varDelta t$ (in red and blue, respectively, left scale) for customers 99 and 630.

Table 2.3. Cardinality and Daily Average Consumption of the New Final Clusters

Final cluster	Cardinality (ratio of new daily profiles)	Average consumption (kWh)
1	4227 (0.01)	51.7
2	12003 (0.05)	45.4
3	2467 (0.03)	34.2
4	6041 (0.32)	31.1
5	18960 (0.58)	17.5
6	190 (0.01)	21
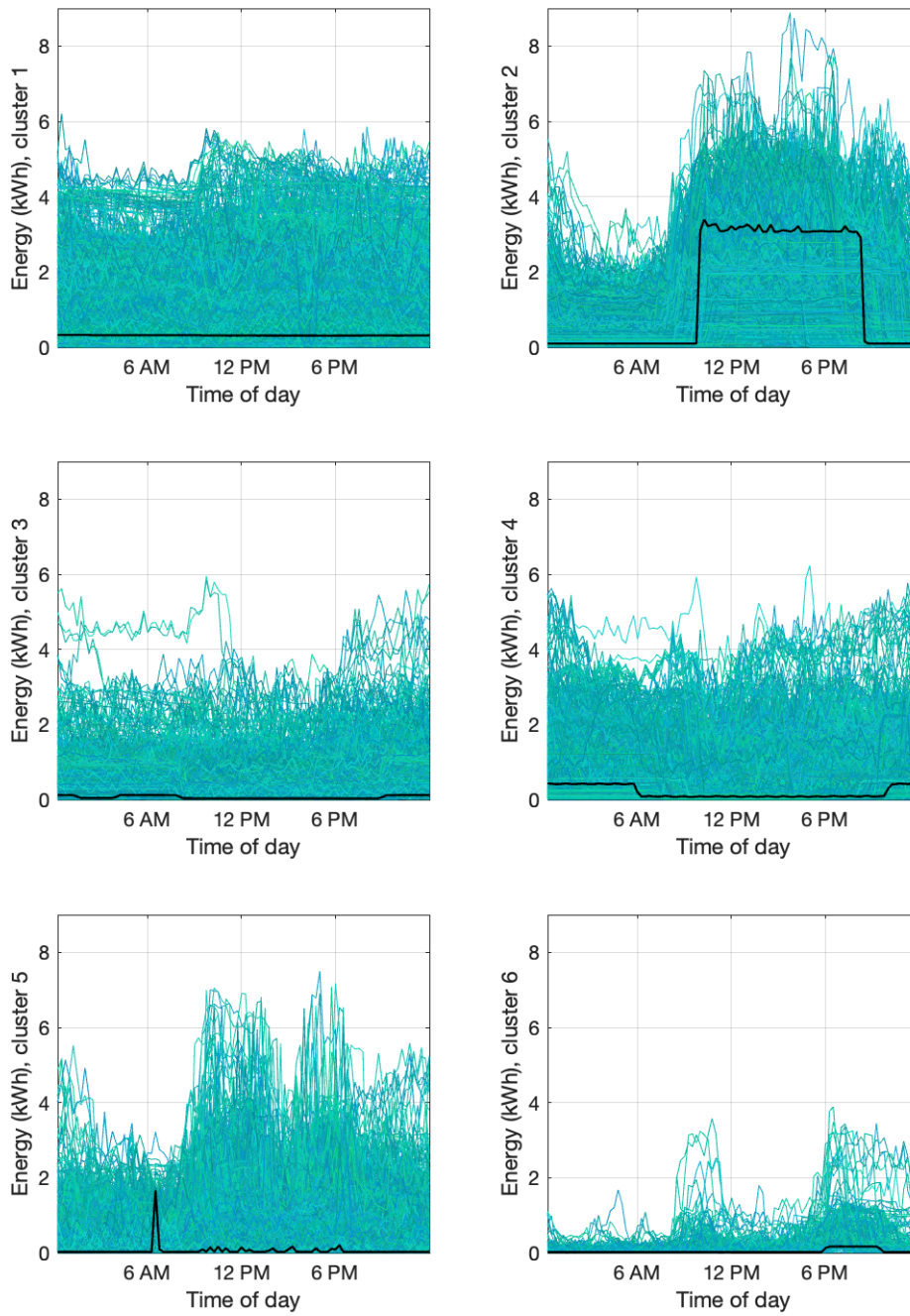
Figure 2.9. Daily consumption profiles of the new final clusters, with cluster centers in black.

Table 2.4. RPCs of Customers and their Combinations

RPC	1	2	3	4	5	6
Number of customers	9	67	242	305	282	20
Combination of RPCs (number of customers of the combination)	2 (5)	4-5 (34)	1-4-5 (6)	1-2-4-5 (135)	1-2-4-5-6 (17)	1-2-3-4-5-6 (20)
	5 (3)	2-5 (27)	2-4-5 (172)	1-3-4-5 (40)	1-2-3-4-5 (252)	
	4 (1)	3-4 (2)	3-4-5 (29)	2-3-4-5 (85)	1-3-4-5-6 (3)	
		1-2 (1)	1-2-4 (7)	2-4-5-6 (24)	2-3-4-5-6 (10)	
		1-4 (1)	1-3-4 (11)	1-2-3-5 (5)		
		1-5 (1)	2-5-6 (7)	1-2-3-4 (12)		
		3-5 (1)	1-2-5 (10)	1-4-5-6 (3)		
				1-2-5-6 (1)		

2.5 Performance Monitoring and Comparison and Sensitive Analysis

Considering the real-world data set, this section first analyzes the monitoring of the clustering algorithm based on the iDB and iXB indices. The comparison with the online algorithm in [29] and the sensitivity analysis employing the number of representatives and the shrink factor are also investigated.

2.5.1 Performance Monitoring of the Online Clustering

To assess the cohesion and separation of clusters produced in the online processing, Figure 2.10 depicts the iDB and iXB indices. A high correlation characterizes the measurement period, and the resulting trend of the values exhibits good performance for the algorithm based on an adequate assignment of objects.

On days when splitting or merging events happen (as shown in Figure 2.4), the indices almost always show a slight increase or decrease, respectively. Conversely, both remain constant when clusters hold from one day to the next. Then, the usefulness of

these incremental forms for monitoring the evolution of the online algorithm is proved, as in other correctly and poorly partitioned data sets [32]. The higher values in the last few days are due to the change in daily patterns attributed to most consumers.
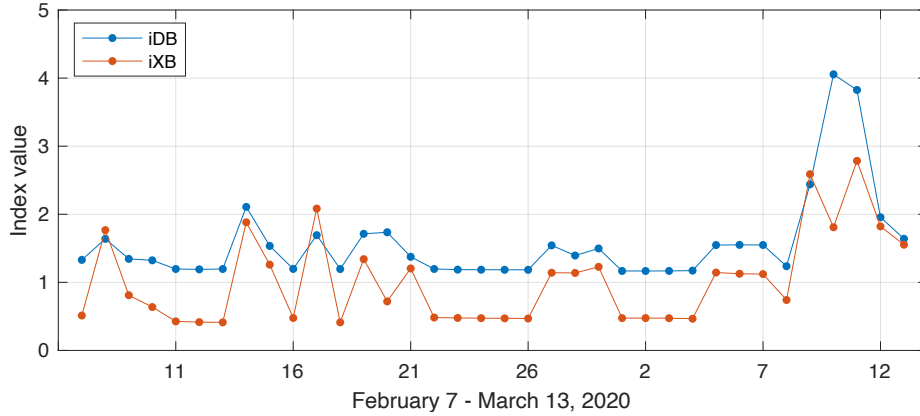
Figure 2.10. Evolution of the iDB and iXB validity indices for the measurement period.

2.5.2 Comparison Analysis

The work uses the algorithm in [29] for comparison purposes since it also processes daily profiles. Concerning the proposed framework, this algorithm has two main differences: 1) it uses centroids, which is not always valid in practice since clusters ideally need to have a spherical structure in the vector space; 2) it discards past daily profiles and preserves this information only through a distance matrix and the centroids themselves, which are updated daily.

The algorithm is applied using the ℓ_2 norm and the first week for the consensus clustering [29], in which the k-means method uses the set $\{2, \dots, 8\}$ for the number of clusters per instance. Table 2.5 gives the result of two cases with different parameter values for the facility cost C^f that decides the split and the minimum distance between centroids d_{min} that favors the merger, where D_p is the mean of the probability distances. The rest of the parameter values comprise the initial number of clusters K^0 obtained from the corresponding dendrogram, the exponential forgetting ν, and the number of disruptive loads γ_{min}, which remain as in the application with electricity data in [29]. Specifically, different parameter values in the second case produce many more splits and mergers.

Table 2.5. Parameter Values and Final Clusters Using the Algorithm in [29]

Case	Parameter value					Final clusters
	K^0	ν	γ_{min}	C^f	d_{min}	
1	7	0.85	5	$6 \times D_p$	$0.5 \times D_p$	6
2	7	0.85	5	$5 \times D_p$	$0.8 \times D_p$	8

Figure 2.11 illustrates the evolution of the iDB and iXB indices in these two cases, where high-magnitude spikes that suggest a poorer assignment of daily profiles are present in both.
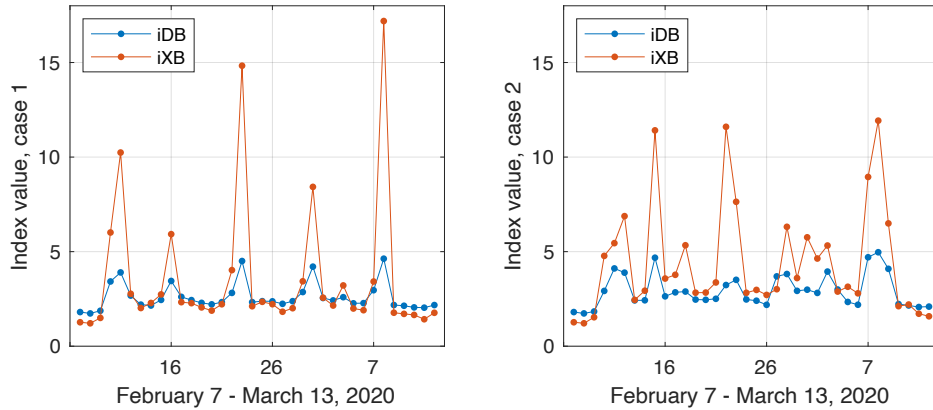
Figure 2.11. Evolution of the iDB and iXB validity indices for the measurement period in two cases with different parameter values using the algorithm in [29].

2.5.3 Sensitivity Analysis

The sensitivity analysis evaluates the influence of changes on the number of representatives N_r and the shrink factor α. Table 2.6 shows solutions for a group of selected values of these parameters.

Considering N_r, the trend is to obtain fewer clusters as this number increases. With one representative (the cluster center itself), nine final clusters result, which is close to the eight final clusters of case 2 using the algorithm in [18] (which employs the centroid). With $N_r = 4$, a high difference between the maximum and minimum values

is obtained for the iDB index. Generally, it is hard to establish the final number of these scattered points in the vector space. With a relatively small number, a large cluster can split incorrectly, but with a large number, two close clusters are more likely to merge. According to the values of the indices in Table 2.6 with the real-world data set, using $N_r \geq 6$ produces adequate results.

Concerning α, it is not practical to take very low or very high values [38]. In addition to $\alpha = 1$, Table 2.6 presents the results with $\alpha = 0.3$ and $\alpha = 0.5$. Similarly, with a smaller value of α, the scattered points shrink little, and merging between two clusters is more likely to occur, whereas larger values cause these points to move more toward the cluster center and favor the splitting.

Table 2.6. Final Clusters and Mean, Standard Deviation, and Difference Between the Maximum and Minimum iDB and iXB Indices for Different Values of N_r and α

N_r	α	Final clusters	Mean, standard deviation, and difference between the maximum and minimum iDB indices	Mean, standard deviation, and difference between the maximum and minimum iXB indices
8	0.4	5	1.57, 0.65, 2.89	1.02, 0.65, 2.37
4	0.4	6	2.92, 5.81, 29.68	1.52, 0.83, 3.27
6	0.4	5	1.47, 0.3, 1.04	1.2, 0.67, 2.23
10	0.4	3	1.27, 0.2, 0.88	0.61, 0.31, 1.36
12	0.4	3	1.19, 0.16, 0.76	0.6, 0.35, 1.34
8	0.3	3	1.41, 0.52, 3.23	1.14, 0.78, 3.5
8	0.5	5	1.28, 0.21, 0.96	0.75, 0.42, 1.52
1	1	9	1.86, 0.19, 0.76	1.34, 0.25, 1.22

2.6 Conclusion

The main idea of this chapter is to propose an online framework for DR characterization. In particular, the DSO can use the proposed framework to obtain and update daily the RPCs and variability of customers and estimate the customer

response to a price signal based on a known RPC, which is suitable for effective energy management on the demand side.

On the other hand, the underlying probability distribution of random deviations in demand from expected values is generally unknown. However, the framework contributes to modeling it since each daily deviation can be considered a realization of the corresponding random variable. Using the parameters of this empirical distribution overcomes the limitation of making distributional assumptions that can result in risky or more conservative costly solutions.

Some technical and practical complications arise with each set **P** of load profiles and the higher amount that needs to be processed daily, in particular: 1) more computational effort is demanded; 2) as clusters extend in the vector space, a larger N_r might eventually be more appropriate for capturing their geometry; 3) the time to find the information about the RPCs of customers and estimate their responses could be higher than that available by the DSO. Thus, a suitable strategy to include in the online algorithm is the well-known sliding window [15], according to which the number of daily profiles remains practically constant, and the information of interest is present. **Algorithm 1** favors the inclusion of this strategy.

While processing and analyzing load data streams represent a significant challenge, the case studies using real-world and simulated daily profiles of Chilean end-users demonstrate the applicability of the proposed approach. According to the results, most consumers use four RPCs within the measurement period. Furthermore, the iDB and iXB indices behavior and the comparison analysis verify the adequate assignment of objects of the online clustering.

Chapter 3

Online Demand Response Scheduling for Renewable Integration Considering Uncertainty

This chapter proposes an online framework for scheduling customers' power responses to support integrating photovoltaic (PV) generation into the distribution system by providing customized control prices. A bi-level optimization model is used to model the interactions between the distribution system operator (DSO) in the upper level, which seeks to maximize its profit, and customers in the lower level, who reduce their electricity bills. The pricing problem, however, is highly challenging since the DSO also requires ensuring the reliable operation of the distribution grid and dealing with uncertainties in consumption and renewable production. Accordingly, the chapter includes and reformulates chance constraints (CCs) for squared nodal voltage and complex power flow in lines. The chance-constrained bi-level problem finally converts to an equivalent mixed-integer second-order cone programming (MISOCP) problem. A case study using real-world local market prices and daily profiles of residential and commercial Chilean end-users on the IEEE-37 node test distribution feeder demonstrates how the proposed framework can schedule demand response (DR) considering uncertainty.

3.1   Motivation

DR is a demand reduction or shift of electricity use by customers to make electricity systems flexible and reliable, which is beneficial under increasing shares of intermittent renewable energy. In a price-based DR program, price signals are formulated to stimulate the customers, encouraging price responsiveness and demand flexibility [35].

In this context, several recent investigations have attempted to determine effective time-varying price signals for customers by exploiting important mathematical programming models. Furthermore, these investigations can be classified based on the designed price signal. On the one hand, some approaches determine a uniform price for a set of dissimilar customers; on the other hand, others determine customized pricing schemes for different groups of customers. Within the former, for example, Tung, Trung, and Bao [42] analyze the energy scheduling of the load-serving entity and formulate its interaction with flexible and inflexible aggregate loads as a bi-level problem that delivers hourly pricing tariffs for the flexible ones. Reference [43] proposes a stochastic bi-level model for transactive energy in which the load-serving entity aims to optimize profit and reserve capacity and similarly designs dynamic prices for the aggregate flexible loads. A hierarchical structure with the DSO, DR providers, and customers is formulated in [44] using a two-stage Stackelberg game, where dynamic and static prices are proposed for the flexible and inflexible end-users, respectively. A stochastic scheduling framework is also designed in [45] to determine the value-based hourly retail prices and the bidding strategy of the distribution company in the day-ahead market. Reference [46] uses reinforcement learning to decide the dynamic retail pricing strategy according to both demand profiles and dissatisfaction levels of customers and market prices. Cai *et al.* [47] present a dynamic pricing model for promoting renewable integration and flattening the grid demand profile using a bi-level optimization that coordinately dispatches flexible loads. In [48], a bi-level model that formulates the interaction between the DSO and end-users allows for designing tariffs with the inclusion of uncertainty in demand, PV generation, and market prices through scenarios. In [35], based on the DR potential of thermostatically controlled loads, a multi-perspective pricing model is proposed to formulate proper DR price signals. Finally, a stochastic bi-level model is proposed in [49] to design daily retail prices for energy management of thermal loads in a community of buildings.

Differently, He, Liu, and Zhang [50] introduce a customized pricing framework based on consumers' occupancy status that comprises a day-ahead stage for designing time-of-use price structures and a real-time stage for final prices using a Stackelberg game. In [51], assuming different types of flexible consumers, a bi-level optimization model is proposed for personalized retail pricing design and demand bidding in the day-ahead market. Finally, a dynamic pricing approach, which includes adaptive customer segmentation and customized demand modeling for clusters, is presented in [52]. However, even though these studies make a significant additional effort by producing dynamic pricing schemes according to customer preferences and providing financial

profits, they do not assess the effect of the final consumption profiles on the distribution grid due to these tariffs. Additionally, the pricing approaches above target a day-ahead programming task, where the price signals are determined based on forecasting future time states. Instead, given the high stochasticity in consumer behavior, renewable generation, and market price, online scheduling is more suitable for handling uncertainty.

This chapter presents an online framework for scheduling customers' power responses to integrate PV production into the distribution system. To this end, the work introduces a bi-level optimization model [2], [53] where the DSO in the upper level is actively involved in the balancing market and maximizes its profit by determining the customized price signals. Customers in the lower level reduce their electricity bills when the prices materialize. Also, the study incorporates a customer characterization stage where the CFSFDP algorithm allows obtaining the RPCs for consumers. The pricing problem, however, is highly challenging since the DSO requires ensuring the reliable operation of the distribution grid and dealing with uncertainties in consumption and PV production. Like studies [54]-[59] in the context of analyzing the stochasticity of distributed resources in the distribution grid, this investigation leverages chance-constrained programming for uncertainty-aware decisions of the DSO. Specifically, CCs for squared voltage and complex power flow are included and reformulated. Therefore, the chance-constrained bi-level problem converts to an equivalent MISOCP problem to find the globally optimal solution. Finally, due to load deviations and forecast errors on PV generation, the chapter presents online scheduling by applying a model predictive control (MPC).

Addressing uncertainty using chance-constrained programming involves, in general, a joint chance-constrained modeling that requires the operational constraints to be satisfied simultaneously, with a joint violation probability ϵ [60]. However, the feasible region induced by this setup is nonconvex and leads to an NP-hard optimization problem. A classical approximation scheme to deal with this intractability is the Bonferroni approximation, which decomposes the joint CC into single CCs, each with a violation probability $\epsilon_i \in [0, \epsilon]$, such that $\sum_i \epsilon_i \leq \epsilon$. Therefore, solving the system of single CCs is significantly straightforward, and any solution satisfying all single constraints also satisfies the joint CC [61]. The shortcoming of the Bonferroni approximation is that the solution quality depends critically on the values of the violation probabilities [62], whose optimal estimation is also an NP-hard problem [61].

Two commonly used choices are the following: 1) an outer approximation that relaxes the requirement of simultaneously satisfying all single constraints and adopts that $\epsilon_i =$

$\epsilon \ \forall i$ (the least conservative choice); 2) an inner approximation where each single CC has the same violation probability, with a value obtained dividing ϵ by the number of single constraints (overly conservative solution) [61], [62]. For applications in power systems, single CCs with outer approximation can be more appropriate as they limit the risk of individual component failures. In practice, this also involves the possibility of using different values of ϵ according to types of constraints, thus pointing to high-risk components or areas [63]. In this investigation, the chance-constrained bi-level model for optimal DR scheduling employs single CCs considering the above criteria.

The main contributions of the chapter are summarized below:

1) An online framework for scheduling customers' power responses to integrate PV production into the distribution system is proposed. The chapter introduces a bi-level optimization model where the DSO in the upper level pursues maximizing its profit by determining the optimal price signals, while customers in the lower level reduce their electricity bills when the prices materialize.

2) The proposed framework ensures the reliable operation of the distribution grid by limiting squared nodal voltages and complex power flows in lines with high probability through CCs. The chance-constrained bi-level problem then converts to an equivalent MISOCP problem.

3) The framework is tested in a case study with real-world market prices and daily profiles of residential and commercial Chilean end-users on the IEEE-37 node test distribution feeder. Results demonstrate how the proposed model can schedule DR in real time. The chapter also explores the impact of uncertainty on system operation and presents a comparison analysis with the deterministic formulation involving the DSO's profit and feasibility of the CCs.

The organization of this chapter is as follows: Section 3.2 provides the problem framework with theoretical foundations and models. Section 3.3 defines the problem formulation under uncertainty. Section 3.4 describes the solution methodology. A case study is presented in Section 3.5 to assess the proposed framework and analyze the uncertainty cost. Finally, Section 3.6 concludes the chapter.

3.2   Problem Framework

This work considers the case in which the DSO aims to schedule demand flexibilities to integrate a lower local PV production and maximize the profit in energy supply. Therefore, the DSO becomes more self-sufficient. To this end, the DSO broadcasts control price signals to customers each hour, which defines the real-time framework of the model. Concerning each customer's decision process, the investigation assumes

that the demand profile derives from an optimizing behavior. Due to the stochastic nature of customer response and PV generation, the DSO requires trading in the balancing market. For simplicity, the chapter only accounts for this energy market for the net difference between consumption and production within the distribution system. However, part of the electrical energy can be procured from other markets in the electricity pool (for example, the day-ahead market) and the futures market [7] to hedge the price uncertainty. The PV production forecast and the locational marginal price (LMP) are also updated each hour of the day. Finally, the framework considers an advanced metering infrastructure that enables two-way communication. Thus, each consumer presents an energy management system that performs optimal scheduling of loads subject to daily consumption preferences to reduce the electricity bill. Figure 3.1 illustrates the problem framework.
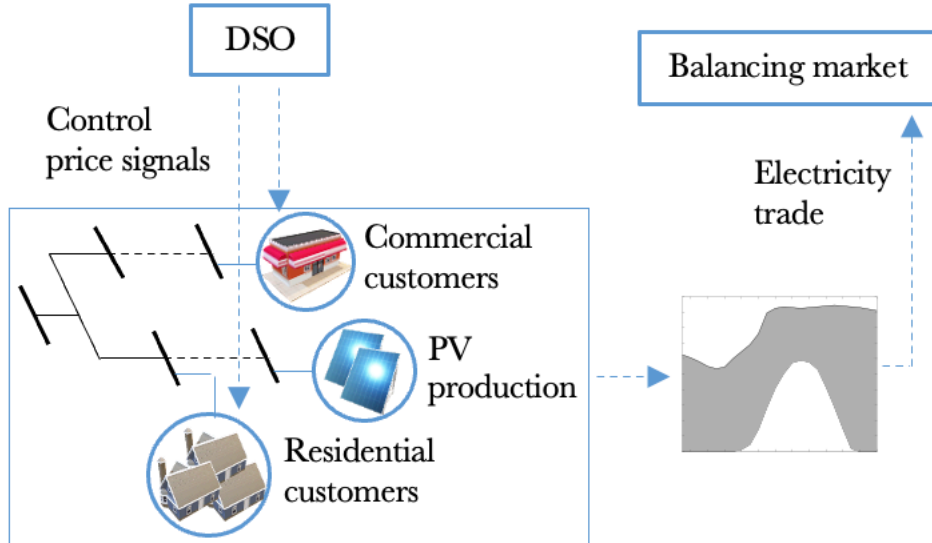
Figure 3.1. Framework for the online DR scheduling.

3.2.1 PV Facility Model

Similarly to the linear model for setting a flexible active power profile in $(2.1) - (2.3)$, let $\bar{s}_{g_t} = \bar{p}_{g_t} + j\bar{q}_{g_t}$ be a complex power forecast to be injected by a PV generation facility $g \in \mathbf{G}$ at time point t, where $\mathbf{G} = \{1, \ldots, G\}$ is the set of PV facilities. Therefore, $\bar{p}_{g_t} = \Re\left(\bar{s}_{g_t}\right)$ is the active component, which represents the active power forecast at the AC side of the inverter, and $\bar{q}_{g_t} = \Im\left(\bar{s}_{g_t}\right)$ the reactive one. Then, $\bar{s}_{g_t} \in \mathcal{S}_{g_t}$,

where $\mathcal{S}_{g_t}$ is the production region that contains the inverter's active and reactive power set points of PV facility g in t, positioned in the complex plane.

The following control strategy for setting a flexible active power profile is defined:

$$0 \leq \bar{p}_{g_t} \leq \bar{p}_{g_t}^{\mathrm{av}}, \quad g \in \mathbf{G}, t \in \mathbf{T}, \qquad\qquad (3.1)$$

Equation (3.1) denotes an active power curtailment provided by the inverter, where $\bar{p}_{g_t}^{\mathrm{av}}$ is the forecast of the available active power, which coincides with the maximum power point and varies stochastically over time based on solar irradiance.

The chapter considers the operation at unit power factor (which means that $\bar{q}_{g_t} = 0$). When solar facilities inject the available active power with this condition, power quality and reliability concerns in the distribution system may emerge for sufficiently high levels of deployed capacities [54]. Therefore, curtailing active power contributes to preventing this. However, the effective coupling of flexible consumption to PV production is also essential to limit power curtailments.

3.2.2 Distribution System Model

This thesis uses the linear *DistFlow* model [64], a lossless approximation of the AC power flow equations, to account for power flows and nodal voltages in the distribution system. The approximate relations facilitate the application of convex reformulations to CCs for a robust solution, which is the focus of the study, rather than providing comprehensive theoretical guarantees for solving the AC optimal power flow (OPF).

Let a radial distribution system comprising the corresponding nodes, collected in the set $\mathbf{N} \cup \{0\}$, $\mathbf{N} = \{1, \dots, N\}$, and distribution lines, represented by the set of pairs of nodes $\mathbf{E} = \{(m, n) : m \in \mathbf{N} \cup \{0\}, n \in \mathbf{N}\}$. Any sending node m lies on the unique path from node 0 to its receiving node n. Node 0 designates the secondary of the power transformer at the substation and is considered the slack node; thus, its nominal voltage v_0 is fixed and known (typically 1.0 pu).

The power injection in slack node 0 depends on the power states of the rest of the nodes. Each node n is characterized at time point t by its complex power $S_{n_t} = P_{n_t} + jQ_{n_t}$, where P_{n_t} and Q_{n_t} denote the active and reactive power, respectively, and the magnitude of its complex voltage v_{n_t}. To provide linear operators, let V_{n_t} be the square of v_{n_t}, with $V_{n_t} \in [V^{\min}, V^{\max}]$, where $V^{\min}$ and $V^{\max}$ are the minimum and

maximum voltage limits. Likewise, each line (m, n) has an impedance value $Z_{mn} = R_{mn} + jX_{mn}$, with R_{mn} and X_{mn} representing the resistance and reactance, respectively. Also, let $S_{mn_t} = P_{mn_t} + jQ_{mn_t}$ designate the complex power flow from node m to node n at time point t, where P_{mn_t} and Q_{mn_t} denote the active and reactive power flows, respectively, and let $S_{mn}^{\max}$ represents the power flow limit.

From the above parameters, the following linear power flow and voltage equations are defined for the model:

$$P_{mn_t} = \mathbf{d}_{mn}{}^\top \mathbf{p^n}_t, \quad (m, n) \in \mathbf{E}, t \in \mathbf{T}, \tag{3.2}$$

$$Q_{mn_t} = \mathbf{d}_{mn}{}^\top \mathbf{q^n}_t, \quad (m, n) \in \mathbf{E}, t \in \mathbf{T}, \tag{3.3}$$

$$V_{n_t} = V_{m_t} - 2\Re\left(Z^*_{mn} S_{mn_t}\right), \quad (m, n) \in \mathbf{E}, t \in \mathbf{T}, \tag{3.4}$$

where $\mathbf{d}_{mn}$ is a vector whose elements correspond to the mnth row of a $[N + 1] \times N$ binary matrix $\mathbf{D}$ that maps the values of active and reactive power in nodes into power flows and voltages, while $\mathbf{p^n}_t = [P_{1_t} \dots P_{N_t}]^\top$ and $\mathbf{q^n}_t = [Q_{1_t} \dots Q_{N_t}]^\top$ are the vectors of active and reactive power of nodes. In $\mathbf{D}$, each element in the mnth row and nth column takes the value 1 if line (m, n) is part of the path from the slack node 0 to node n and the value 0 otherwise.

The expressions for the (net) active and reactive power in each node are as follows:

$$P_{n_t} = \mathbf{h}_n{}^\top \mathbf{p}_t, \quad n \in \mathbf{N}, t \in \mathbf{T}, \tag{3.5}$$

$$Q_{n_t} = \mathbf{h}_n{}^\top \mathbf{q}_t, \quad n \in \mathbf{N}, t \in \mathbf{T}, \tag{3.6}$$

where $\mathbf{h}_n$ is a vector, whose elements correspond to the nth column of a $[L + G] \times N$ binary matrix $\mathbf{H}$ that indicates the belonging of each customer l and PV facility g to the corresponding node n, and $\mathbf{p}_t$ and $\mathbf{q}_t$ are vectors comprising respectively active and reactive power of both customers and solar facilities at time point t.

For exposition simplicity, the investigation assumes a balanced distribution system.

3.3 Chance-constrained Problem Formulation

3.3.1 Uncertainty Modeling

Both customer consumption and PV production are random variables that take any value within specific ranges; hence, continuous probability distributions can describe them. The modeling of their corresponding active power values can be of the form:

$$p_{l_t} = \bar{p}_{l_t} + \xi_{l_t}, \quad l \in \mathbf{L}, t \in \mathbf{T}, \tag{3.7}$$

$$p_{g_t} = \bar{p}_{g_t} + \xi_{g_t}, \quad g \in \mathbf{G}, t \in \mathbf{T}, \tag{3.8}$$

where $\xi_{l_t}, \xi_{g_t} \in \mathbb{R}$ are stochastic and represent the deviation of the customer response and the PV forecast error, respectively.

Thus, for each time point t, a random vector $\boldsymbol{\xi}_t \in \mathbb{R}^N$ that collects the resulting active power deviation and forecast error in the nodes can be determined. Also, it is possible to express the power flows in lines and nodal voltages in terms of $\boldsymbol{\xi}_t$. For the first case:

$$P_{mn_t} = \bar{P}_{mn_t} + \mathbf{d}_{mn}^\top \boldsymbol{\xi}_t, \quad (m,n) \in \mathbf{E}, t \in \mathbf{T}, \tag{3.9}$$

$$Q_{mn_t} = \bar{Q}_{mn_t} + \mathbf{d}_{mn}^\top \boldsymbol{\Theta}_t \boldsymbol{\xi}_t, \quad (m,n) \in \mathbf{E}, t \in \mathbf{T}, \tag{3.10}$$

where $\bar{P}_{mn_t}$ and $\bar{Q}_{mn_t}$ are computed using (3.2) and (3.3) with expected and forecast values, and $\boldsymbol{\Theta}_t$ is an $N \times N$ diagonal matrix with the tangents correspondingly relating the active and reactive nodal deviations or errors at time point t.

Analogously, the uncertain nodal voltages are:

$$V_{n_t} = \bar{V}_{n_t} - 2\mathbf{d}_n^\top (\mathbf{RD} + \mathbf{XD}\boldsymbol{\Theta}_t)\boldsymbol{\xi}_t, \quad n \in \mathbf{N}, t \in \mathbf{T}, \tag{3.11}$$

where $\bar{V}_{n_t}$ is computed using (3.4) with expected and forecast values, $\mathbf{d}_n$ is a vector whose elements correspond to the nth column of $\mathbf{D}$, and $\mathbf{R}$ and $\mathbf{X}$ are $N \times N$ diagonal matrices with the resistance and reactance values of lines, respectively.

3.3.2 Chance-Constrained Bi-Level Formulation

The thesis presents a hierarchical optimization structure that comprises two levels. In the upper, the DSO determines the price signals under uncertainty, anticipating the customers' response, while in the lower, each adjusts its consumption independently to reduce the electricity bill. The formulation of the chance-constrained bi-level model is as follows:

$$\max_{\lambda_{l_t}} \mathbb{E}\left[\sum_{t=1}^{T}\sum_{l=1}^{L}\lambda_{l_t}p_{l_t}\Delta t - \sum_{t=1}^{T}\lambda_t^b\left(\sum_{l=1}^{L}p_{l_t} - \sum_{g=1}^{G}p_{g_t}\right)\Delta t\right] \tag{3.12}$$

$$\text{s.t. } \lambda^{\min} \leq \lambda_{l_t} \leq \lambda^{\max}, \quad l \in \mathbf{L}, t \in \mathbf{T}, \tag{3.13}$$

$$\frac{1}{T}\sum_{t=1}^{T}\lambda_{l_t} \leq \lambda^r, \quad l \in \mathbf{L}, \tag{3.14}$$

$$(3.1) - (3.6),$$

$$\mathbb{P}\{V^{\min} \leq V_{n_t} \leq V^{\max}\} \geq 1 - \epsilon_{\mathrm{v}}, \quad n \in \mathbf{N}, t \in \mathbf{T}, \tag{3.15}$$

$$\mathbb{P}\{P_{mn_t}{}^2 + Q_{mn_t}{}^2 \leq S_{mn}^{\max 2}\} \geq 1 - \epsilon_{\mathrm{i}}, \quad (m,n) \in \mathbf{E}, t \in \mathbf{T}, \tag{3.16}$$

$$\bar{p}_{l_t} \in \arg\min_{\bar{p}_{l_t}}\left[\sum_{t=1}^{T}\lambda_{l_t}\bar{p}_{l_t}\Delta t\right], \quad l \in \mathbf{L}, \tag{3.17}$$

$$\text{s.t. } (2.1) - (2.3)$$

In the upper-level problem, the objective function (3.12) is maximized under the expected value over the probability distribution of uncertainty. The objective function represents the difference between revenue and cost (the profit) for the DSO by supplying electricity to customers. Each λ_{l_t} is a decision price, and λ_t^b is the LMP. At time point t, λ_t^b is known from the balancing market clearing and forecasted for period $t+1$ to T. Equation (3.13) defines a feasible region for the decision prices between a minimum $\lambda^{\min}$ and a maximum $\lambda^{\max}$ value. In (3.14), the mean of prices λ_{l_t} over the time horizon T does not surpass the regulated price λ^r, which implies that consumers do not incur financial losses beyond λ^r. The outer approximation set of CCs in (3.15) and (3.16) ensures that in the schedule of DR, technical constraints for magnitudes of squared nodal voltages (with linear inequalities) and complex power

flows in lines (with quadratic inequalities) do not violate their limits with probabilities $1 - \epsilon_v$ and $1 - \epsilon_i$, where ϵ_v and ϵ_i are the violation probabilities for voltages and power flows, respectively. Lastly, (3.17) represents the lower-level problems, which enforce the expected optimal consumption of customers given the decision variables λ_{l_t}.

Therefore, control prices guarantee that power consumption and generation have a feasible power flow solution satisfying the constraints on voltages and power flows.

3.4 Solution Methodology

Due to (3.15) and (3.16), the resolution of the problem is computationally intractable. The section first provides second-order cone (SOC) reformulations for these CCs by assuming the DSO knows the probability density function of vectors $\boldsymbol{\xi}_t$, specifically that each $\boldsymbol{\xi}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t)$, where $\boldsymbol{\Sigma}_t \in \mathbb{R}^{N \times N}$ is diagonal. Reference [9], for instance, discusses several attractive properties that support the selection of the normal distribution. Furthermore, due to (3.17), tractability depends on reformulating the problem into a single-level one. Since lower-level problems are linear in the decision variables $\bar{p}_{l_t}$, and control prices λ_{l_t} are parameters, their corresponding Karush-Kuhn-Tucker (KKT) optimality conditions [66] can replace them, which are both necessary and sufficient for global optimality. Finally, the objective function includes bilinear products that produce nonlinearity; however, linearization is possible using duality theory [66].

3.4.1 Reformulation of the CCs

1) Voltage CCs: Every single constraint in (3.15) with a limit for the magnitude of squared nodal voltage is a CC of the general form:

$$\mathbb{P}\left\{a\left(\lambda_{l_t}\right) + \mathbf{b}\left(\lambda_{l_t}\right)^\top \boldsymbol{\xi}_t \leq c\right\} \geq 1 - \epsilon, \qquad (3.18)$$

where $a\left(\lambda_{l_t}\right) \in \mathbb{R}$ and $\mathbf{b}\left(\lambda_{l_t}\right) \in \mathbb{R}^N$ denote affine functions of the decision variable λ_{l_t} and describe the nominal squared voltage and the influence of $\boldsymbol{\xi}_t$ on the constraint, respectively, while c is a constant representing the voltage limit.

Equation (3.18) is convex if, in general, $\boldsymbol{\xi}_t$ is an elliptical log-concave distribution function (which includes the considered normal). In particular, the left-hand side is a random variable ξ whose distribution represents the variations in the corresponding

voltage, with mean $a\left(\lambda_{l_t}\right)$ and standard deviation $\left\|\mathbf{b}\left(\lambda_{l_t}\right)^{\top}\mathbf{\Sigma}_t^{1/2}\right\|_2$. Therefore, the CC is reformulated exactly into the following SOC constraint (see Appendix B):

$$a\left(\lambda_{l_t}\right) \leq c - \Phi^{-1}(1 - \epsilon)\left\|\mathbf{b}\left(\lambda_{l_t}\right)^{\top}\mathbf{\Sigma}_t^{1/2}\right\|_2, \qquad (3.19)$$

where Φ is the cumulative distribution function of the standard normal distribution, and because $\epsilon < {}^1\!/_2$, $\Phi^{-1}(1 - \epsilon) > 0$.

In (3.19), the term $a\left(\lambda_{l_t}\right) \leq c$ represents the nominal constraint, that is, the constraint resulting when neglecting uncertainty. The second term $\Phi^{-1}(1 - \epsilon)\left\|\mathbf{b}\left(\lambda_{l_t}\right)^{\top}\mathbf{\Sigma}_t^{1/2}\right\|_2$ represents the adjustment of the nominal available capacity, necessary to ensure the system against errors and deviations. This term is always negative, which means it always leads to a reduction in the available capacity. This reduction can be interpreted as a security margin against uncertainty, that is, an uncertainty margin [9].

Based on the above reformulation, CCs of squared nodal voltage magnitudes in (3.15) recast as:

$$\bar{V}_{n_t} \leq V^{\max} - 2\Phi^{-1}(1 - \epsilon_{\mathrm{v}})\left\|\mathbf{d}_n^{\top}[\mathbf{RD} + \mathbf{XD\Theta}_t]\mathbf{\Sigma}_t^{1/2}\right\|_2, \quad n \in \mathbf{N}, t \in \mathbf{T}, \qquad (3.20)$$

$$-\bar{V}_{n_t} \leq -V^{\min} - 2\Phi^{-1}(1 - \epsilon_{\mathrm{v}})\left\|\mathbf{d}_n^{\top}[\mathbf{RD} + \mathbf{XD\Theta}_t]\mathbf{\Sigma}_t^{1/2}\right\|_2, \quad n \in \mathbf{N}, t$$
$$\in \mathbf{T}, \qquad (3.21)$$

2) Power Flow CCs: Every single constraint in (3.16) with a limit for the magnitude of complex power flow is a CC of the general form:

$$\mathbb{P}\left\{\left[a_1\left(\lambda_{l_t}\right) + \mathbf{b}_1\left(\lambda_{l_t}\right)^{\top}\mathbf{\xi}_t\right]^2 + \left[a_2\left(\lambda_{l_t}\right) + \mathbf{b}_2\left(\lambda_{l_t}\right)^{\top}\mathbf{\xi}_t\right]^2 \leq c^2\right\} \geq 1 - \epsilon, \qquad (3.22)$$

where the squared random variables on the left-hand side ξ_1 and ξ_2 represent the operating points of the active and reactive power flows in the line contained in the feasible region $\mathcal{S} \in \mathbb{R}^2$, while c is a constant representing the apparent power limit.

From (3.22), $\mathcal{S}$ is convex and can be inner approximated by the I-sided polygon $\mathcal{P}$ inscribed inside $\mathcal{S}$, with I even and selected in advance [65]. Then, the work defines the following formulation: $\mathcal{P} = \{\mathbf{Ws} \leq \mathbf{0}\}$, where $\{\mathbf{Ws} \leq \mathbf{0}\}$ is a set of half-space linear constraints, $\mathbf{W}$ is an $I \times 3$ matrix of coefficients, and $\mathbf{s} \in \mathbb{R}^3$ is the vector formed by each of the elements of (3.22).

The following linear constraints are specified to find the coefficients of $\mathbf{W}$:

$$\left(\xi_2 - \xi_{2_i}\right) - \eta_i\left(\xi_1 - \xi_{1_i}\right) \leq 0, \quad i = 1, \dots, {}^I\!/_2, \qquad (3.23)$$

$$-\left(\xi_2 - \xi_{2_i}\right) + \eta_i\left(\xi_1 - \xi_{1_i}\right) \leq 0, \quad i = {}^I\!/_2 + 1, \dots, I, \quad (3.24)$$

where η_i is the slope of side i, while ξ_{1_i} and ξ_{2_i} are vertices of $\mathcal{P}$, respectively defined as $\xi_{1_i} = c\cos[(i-1)\theta]$ and $\xi_{2_i} = c\sin[(i-1)\theta]$, with $\theta = {}^{2\pi}\!/_I$ being the angle between two consecutive vertices.

Substituting these definitions and rearranging conveniently, (3.23) and (3.24) result as:

$$-\eta_i\xi_1 + \xi_2 + (\eta_i\cos[(i-1)\theta] - \sin[(i-1)\theta])c \leq 0, \quad i = 1, \dots, {}^I\!/_2, \quad (3.25)$$

$$\eta_i\xi_1 - \xi_2 + (\sin[(i-1)\theta] - \eta_i\cos[(i-1)\theta])c \leq 0, \quad i = {}^I\!/_2 + 1, \dots, I, \quad (3.26)$$

which together imply the matrix product $\mathbf{Ws}$, and from which the coefficients of $\mathbf{W}$ are derived with $\eta_i = (\sin[(i)\theta] - \sin[(i-1)\theta])/(\cos[(i)\theta] - \cos[(i-1)\theta])$.

By expressing as $w_{i,1}$, $w_{i,2}$, and $w_{i,3}$ the coefficients of any ith row of $\mathbf{W}$, the following linear CC is generalized:

$$\mathbb{P}\left\{w_{i,1}\left[a_1(\lambda_{l_t}) + \mathbf{b}_1(\lambda_{l_t})^\top\boldsymbol{\xi}_t\right] + w_{i,2}\left[a_2(\lambda_{l_t}) + \mathbf{b}_2(\lambda_{l_t})^\top\boldsymbol{\xi}_t\right] + w_{i,3}c \leq 0\right\}$$
$$\geq 1 - \epsilon, \quad i = 1, \dots, I, \qquad (3.27)$$

CCs in (3.27) can be converted into one with the same structure as (3.18) due to $\boldsymbol{\xi}_t$ in both random variables. Thus, the following set of constraints for the half-space approximation of $\mathcal{S}$ can replace the CCs of complex power flows in (3.16):

$$w_{i,1}\bar{P}_{mn_t} + w_{i,2}\bar{Q}_{mn_t}$$
$$\leq -w_{i,3}S_{mn}^{\max} - \Phi^{-1}(1-\epsilon_i)\left\|\left[w_{i,1}\mathbf{d}_{mn}^{\top} + w_{i,2}\mathbf{d}_{mn}^{\top}\mathbf{\Theta}_t\right]\mathbf{\Sigma}_t^{1/2}\right\|_2, \quad i$$
$$= 1,\dots,I, (m,n) \in \mathbf{E}, t \in \mathbf{T}, \tag{3.28}$$

3.4.2 Reformulation of the Lower-Level Problems

Applying the KKT formulation to the lower-level problems gives the following expressions:

$$\Delta t \lambda_{l_1} - \beta_{l_1}^{\min} + \beta_{l_1}^{\max} + \psi_{l_2}^{\mathrm{d}} - \psi_{l_2}^{\mathrm{u}} - \varepsilon_l \Delta t = 0, \quad l \in \boldsymbol{L}, \tag{3.29}$$

$$\Delta t \lambda_{l_t} - \beta_{l_t}^{\min} + \beta_{l_t}^{\max} - \psi_{l_t}^{\mathrm{d}} + \psi_{l_{t+1}}^{\mathrm{d}} + \psi_{l_t}^{\mathrm{u}} - \psi_{l_{t+1}}^{\mathrm{u}} - \varepsilon_l \Delta t = 0, \quad l \in \mathbf{L}, t$$
$$= 2,\dots,T-1, \tag{3.30}$$

$$\Delta t \lambda_{l_T} - \beta_{l_T}^{\min} + \beta_{l_T}^{\max} - \psi_{l_T}^{\mathrm{d}} + \psi_{l_T}^{\mathrm{u}} - \varepsilon_l \Delta t = 0, \quad l \in \mathbf{L}, \tag{3.31}$$

$$0 \leq \beta_{l_t}^{\min} \perp \bar{p}_{l_t} - p_{l_t}^{\min} \geq 0, \quad l \in \mathbf{L}, t \in \mathbf{T}, \tag{3.32}$$

$$0 \leq \beta_{l_t}^{\max} \perp -\bar{p}_{l_t} + p_{l_t}^{\max} \geq 0, \quad l \in \mathbf{L}, t \in \mathbf{T}, \tag{3.33}$$

$$0 \leq \psi_{l_t}^{\mathrm{d}} \perp \bar{p}_{l_t} - \bar{p}_{l_{t-1}} + r_{l_t}^{\mathrm{d}} \Delta t \geq 0, \quad l \in \mathbf{L}, t = 2,\dots,T, \tag{3.34}$$

$$0 \leq \psi_{l_t}^{\mathrm{u}} \perp -\bar{p}_{l_t} + \bar{p}_{l_{t-1}} + r_{l_t}^{\mathrm{u}} \Delta t \geq 0, \quad l \in \mathbf{L}, t = 2,\dots,T, \tag{3.35}$$

$$0 \leq \varepsilon_l \perp \sum_{t=1}^{T} \bar{p}_{l_t} \Delta t - e_l \geq 0, \quad l \in \mathbf{L}, \tag{3.36}$$

which include in $(3.29) - (3.31)$ the gradient of the Lagrange function concerning variables $\bar{p}_{l_t}$ (different for cases $t = 1$ and $t = T$ due to ramp constraints), and the complementary-slackness conditions between the non-negative dual variables and the inequality constraints of the primal problem.

Because of the nonlinearity of complementary-slackness conditions, the study uses the mixed-integer method of [67], resulting in the following set of linear constraints:

$$0 \leq \beta_{l_t}^{\min} \leq M^{\mathrm{d}} B_{l_t}^{\min}, \quad l \in \mathbf{L}, t \in \mathbf{T}, \tag{3.37}$$

$$0 \leq \bar{p}_{l_t} - p_{l_t}^{\min} \leq M^{\mathrm{p}}\left(1 - B_{l_t}^{\min}\right), \quad l \in \mathbf{L}, t \in \mathbf{T}, \tag{3.38}$$

$$0 \leq \beta_{l_t}^{\max} \leq M^{\mathrm{d}} B_{l_t}^{\max}, \quad l \in \mathbf{L}, t \in \mathbf{T}, \tag{3.39}$$

$$0 \leq -\bar{p}_{l_t} + p_{l_t}^{\max} \leq M^{\mathrm{p}}\big(1 - B_{l_t}^{\max}\big), \quad l \in \mathbf{L}, t \in \mathbf{T}, \tag{3.40}$$

$$0 \leq \psi_{l_t}^{\mathrm{d}} \leq M^{\mathrm{d}} \Psi_{l_t}^{\mathrm{d}}, \quad l \in \mathbf{L}, t = 2, \dots, T, \tag{3.41}$$

$$0 \leq \bar{p}_{l_t} - \bar{p}_{l_{t-1}} + r_{l_t}^{\mathrm{d}} \Delta t \leq M^{\mathrm{p}}\big(1 - \Psi_{l_t}^{\mathrm{d}}\big), \quad l \in \mathbf{L}, t = 2, \dots, T, \tag{3.42}$$

$$0 \leq \psi_{l_t}^{\mathrm{u}} \leq M^{\mathrm{d}} \Psi_{l_t}^{\mathrm{u}}, \quad l \in \mathbf{L}, t = 2, \dots, T, \tag{3.43}$$

$$0 \leq -\bar{p}_{l_t} + \bar{p}_{l_{t-1}} + r_{l_t}^{\mathrm{u}} \Delta t \leq M^{\mathrm{p}}\big(1 - \Psi_{l_t}^{\mathrm{u}}\big), \quad l \in \mathbf{L}, t = 2, \dots, T, \tag{3.44}$$

$$0 \leq \varepsilon_l \leq M^{\mathrm{d}} E_l, \quad l \in \mathbf{L}, \tag{3.45}$$

$$0 \leq \sum_{t=1}^{T} \bar{p}_{l_t} \Delta t - e_l \leq M^{\mathrm{p}}(1 - E_l), \quad l \in \mathbf{L}, \tag{3.46}$$

$$B_{l_t}^{\min}, B_{l_t}^{\max} \in \{0,1\}, \quad l \in \mathbf{L}, t \in \mathbf{T}, \tag{3.47}$$

$$\Psi_{l_t}^{\mathrm{d}}, \Psi_{l_t}^{\mathrm{u}} \in \{0,1\}, \quad l \in \mathbf{L}, t = 2, \dots, T, \tag{3.48}$$

$$E_l \in \{0,1\}, \quad l \in \mathbf{L}, \tag{3.49}$$

where M^{p} and M^{d} are large enough positive constants valid as upper bounds for the primal and dual variables.

3.4.3 Reformulation of the Objective Function

Bilinear products $\lambda_{l_t} \bar{p}_{l_t} \Delta t$ in the objective function produce nonlinearity. However, based on the strong-duality property, lower-level problems (with the same form of bilinear products) and their corresponding dual can be interchangeable since the optimal objective function value is equal in the primal and dual problems. Using the strong-duality equality results in (3.50). Therefore, the chance-constrained bi-level problem converts to an equivalent MISOCP problem, where the globally optimal solution is obtained by applying the branch and bound algorithm [66].

$$\max_{\lambda_{l_t}} \mathbb{E}\left\{ \sum_{l=1}^{L} \left[\sum_{t=1}^{T} \big(p_{l_t}^{\min} \beta_{l_t}^{\min} - p_{l_t}^{\max} \beta_{l_t}^{\max}\big) + \sum_{t=2}^{T} \big(-r_{l_t}^{\mathrm{d}} \Delta t \psi_{l_t}^{\mathrm{d}} - r_{l_t}^{\mathrm{u}} \Delta t \psi_{l_t}^{\mathrm{u}}\big) + e_l \varepsilon_l \right] \right.$$
$$\left. - \sum_{t=1}^{T} \lambda_t^b \left(\sum_{l=1}^{L} p_{l_t} - \sum_{g=1}^{G} p_{g_t} \right) \Delta t \right\} \tag{3.50}$$

s.t. $(3.13) - (3.14), (3.1) - (3.6), (3.20) - (3.21), (3.28), (3.29)$
$\qquad - (3.31), (3.37) - (3.49)$

3.4.4 MPC-Based Implementation

Due to the presence of uncertainty, the MPC [2] can be suitable for this decision-making problem, as it enables the model (3.50) to achieve optimal prices for the current and future time points based on newly available information of the system states within the horizon T; in particular, information concerning to forecasts on PV facilities and prices λ_t^b from the balancing market. A similar implementation is present in [54]. Hence, instead of $t = 1$ in (3.50), t could be $1, \dots, T$. The MPC involves the steps outlined in **Algorithm 2**, where the look-ahead time decreases each day, being T the final time point at any t.

Algorithm 2 Online DR Scheduling

Input: Problem data, daily RPC of each customer

1:   $t = 1$;
2:   **while** $t <= T$, **do**
3:       Update the forecasts of available active power $\bar{p}_{g_t}^{\mathrm{av}}$ and
         balancing market price λ_t^b, $t{:}T$;
4:       Solve (3.50);
5:       Implement optimal solutions $[\lambda_{l_t} \dots \lambda_{l_T}]^{\mathsf{T}}$;
6:       Save the current optimal values of prices λ_{l_t} and the power
         responses p_{l_t} and q_{l_t} of each customer;
7:       $t \leftarrow t + 1$;
8:   **end while**

3.5 Case Study

This section presents a case study to demonstrate the benefits of the approach. First, a characterization stage using electricity data of residential and commercial Chilean end-users allows for determining the RPCs of these customers and their combinations. Then, the section shows an application example for DR scheduling and investigates how CCs impact the solution of the scheduling problem. In the algorithmic implementation, the work employs the toolbox YALMIP [68] on MATLAB 2022 with

the Gurobi optimization solver, running on a MacBook Air with a processor M1 and memory of 8 GB.

3.5.1 System Setup

The case study uses a modified version of the IEEE-37 node test feeder [69], replacing original spot loads with SM data of 125 end-users. Figure 3.2 shows the distribution system, which operates at 4.8 kV line-to-line nominal voltage. The data correspond to the three weeks of February 22 to March 13, 2020. In addition, two PV facilities are added to the distribution grid to emulate the renewable penetration at the DSO scale. Table 3.1 summarizes the allocation of consumers in the load nodes and the nodes for PV production. The test day considered is March 14, for which the study uses the LMP associated with the substation. The profiles of the PV generation are simulated based on historical weather data [70] at the geographical area of end-users.
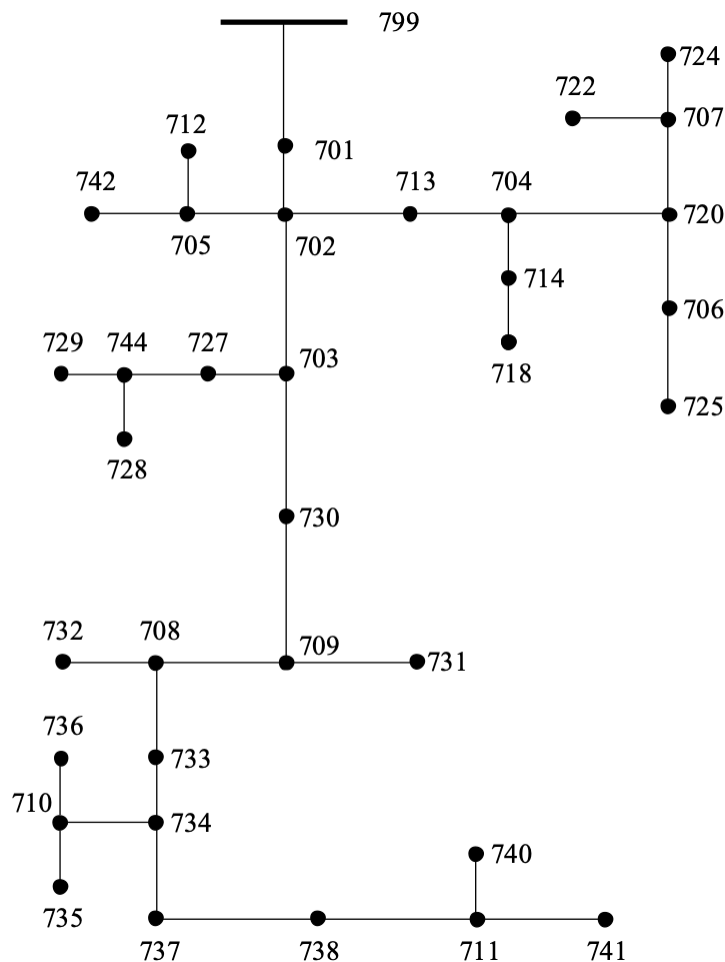

Figure 3.2. Modified IEEE-37 node test feeder considered in the case study.

Table 3.1. Number of Customers and PV Facilities in Nodes of the Test Feeder

Load node	Number of customers	Generation node	Number of PV facilities
701	37	702	1
712	1	709	1
713	1		
718	1		
720	2		
722	4		
724	1		
725	1		
727	1		
728	3		
729	1		
730	2		
731	2		
732	1		
733	3		
734	1		
735	3		
736	1		
737	5		
738	5		
740	4		
741	43		
742	1		
744	1		

The work also defines a feasible region of control prices between 40 \$/MWh and 80 \$/MWh, with a regulated price of 70 \$/MWh. The violation probabilities in CCs are $\epsilon_v = 0.05$ and $\epsilon_i = 0.01$. Limits for nodal voltages are set to 1.05 pu and 0.95 pu, whereas for power flows, their calculation employs the line-to-line nominal voltage and ampacity of electrical conductors [71]. The upper bounds in the lower-level problems are $M^p = 100$ and $M^d = 1000$, while $I = 12$ for polygon $\mathcal{P}$.

Finally, to construct vectors ξ_t, 90 samples are drawn from normal distributions of nodal demand deviations and forecast errors. The corresponding standard deviation is assumed to be 3 % of the expected or forecast active power for the first and second hours, 6 % for hours three to five, 10 % for hours six to nine, 15 % for hours ten to 14, and 20 % for the rest of future time points. The nodal expected or forecast value considers the aggregate average demand or the available PV generation. Power factors of 0.93 and 0.86 are assumed for residential and commercial end-users, respectively. Then, the computation of elements in Θ_t comprises the average power factor for load nodes and the unit power factor for generation nodes.

3.5.2 Customer Characterization and Modeling

The application of the CFSFDP to the set of daily profiles allows the identification of three clusters through the plots of minimum distance δ as a function of local density ρ for each object and their products in decreasing order. Figure 3.3 illustrates the result and highlights the selected centers with colored and bigger dots. One (with the violet color) is remarkably different because of its higher neighborhood compared to the other two.
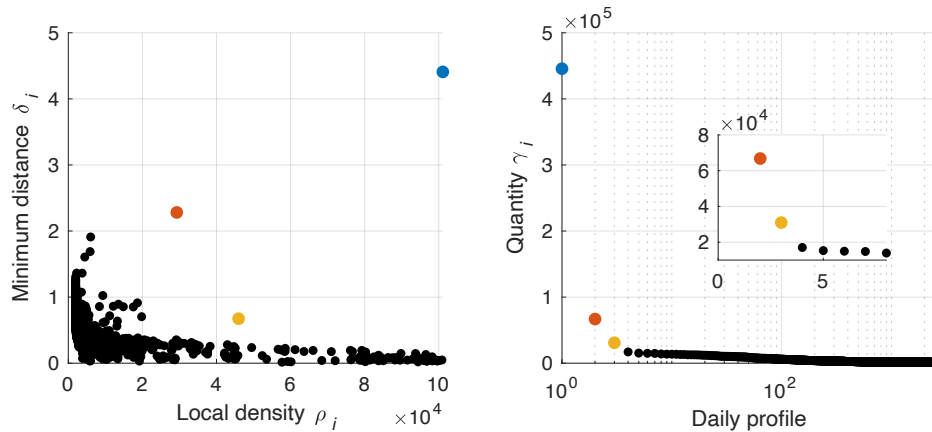

Figure 3.3. Plots for identification of clusters: the decision graph and quantities γ in decreasing order (including a zoom to better distinguish the last two).

The daily profiles of the clusters are depicted in Figure 3.4, with cardinalities of 421, 459, and 1740, respectively. Cluster 1 has profiles with a more uniform trend in consumption. Instead, the consumption in cluster 2 concentrates during working hours, which infers that this behavior includes mainly commercial establishments and businesses. Finally, cluster 3 exhibits a higher consumption in the afternoon and night. Specifically, cluster centers 1, 2, and 3 correspond to the violet, brown, and green dots in Figure 3.3.
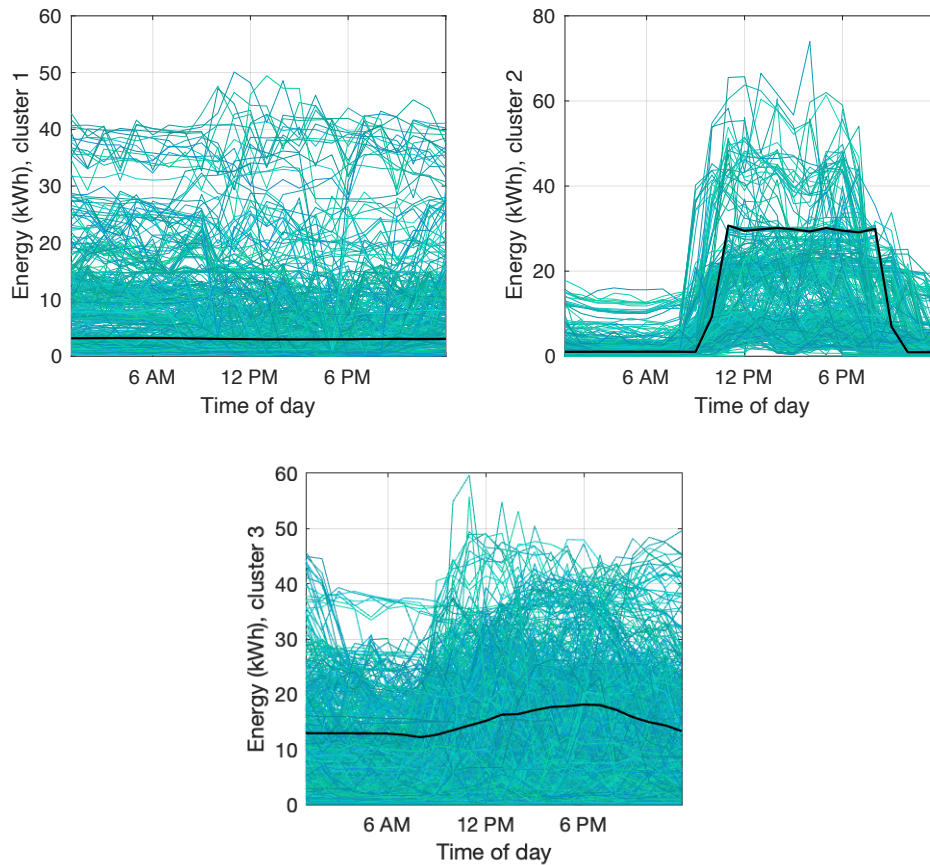

Figure 3.4. Daily consumption profiles of clusters, with cluster centers in black.

Table 3.2 summarizes the RPCs and the combinations of these three daily patterns. Most customers use two RPCs during the period, mainly consumption patterns 1 and 3.

For the test day, the study assumes that consumers use the RPC with the highest probability for that day, which results in 26 customers using the RPC 1, 11 the RPC 2, and 88 the RPC 3. Figure 3.5 depicts the customers' aggregate average demand

from their corresponding RPC, the available PV generation profile, and the market prices for the test day.

Table 3.2. RPCs of Customers and their Combinations

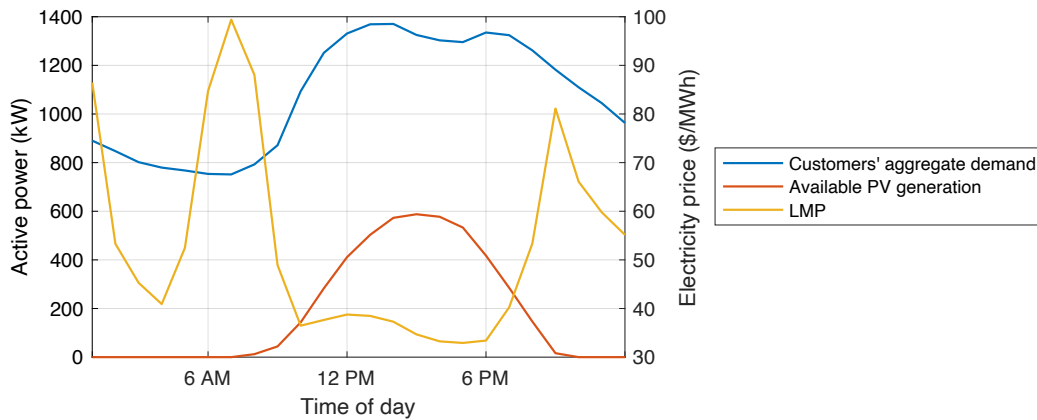| RPC	1	2	3
Number of customers	31	67	27
Combination of RPCs (Number of customers of the combination)	1 (1) 2 (3) 3 (27)	1-2 (3) 1-3 (46) 2-3 (18)	1-2-3 (27)

Figure 3.5. Expected customers' aggregate average demand, available PV generation, and LMP on March 14, 2020.

3.5.3 DR Scheduling

With the RPCs of customers, **Algorithm 2** can deliver optimal control prices that maximize DSO's profit in energy supply. For conciseness, the chapter presents the results for time point $t = 1$, which is the most problematic since it implicates solving the scheduling problem considering all time points of the day with their corresponding uncertainties.

Figure 3.6 illustrates the three control price signals to be broadcasted by the DSO according to the highest probability RPC on the test day. For the first, the lowest price of 64 $/MWh is uniform from 10 AM, comprising the higher PV production and the lower market prices. Therefore, customers with RPC 1 are encouraged to change their

consumption by shifting some activities to this period. A different situation characterizes the second control signal, with the lowest price of 40 \$/MWh ranging from 3 AM to 8 AM, involving only the low market prices of the early morning. One possible reason could be that a coincidence in consumption with the rest of the end-users can produce a high-demand period, leading to voltage violations beyond the acceptable limit. The third control signal presents the regulated price for the entire day. In general, customers with RPC 3 exhibit a higher consumption in the hours of higher PV production and lower market prices, which is of the DSO's interest. In addition, the DSO expects this RPC for most consumers; thus, it maximizes its revenue by broadcasting this price.
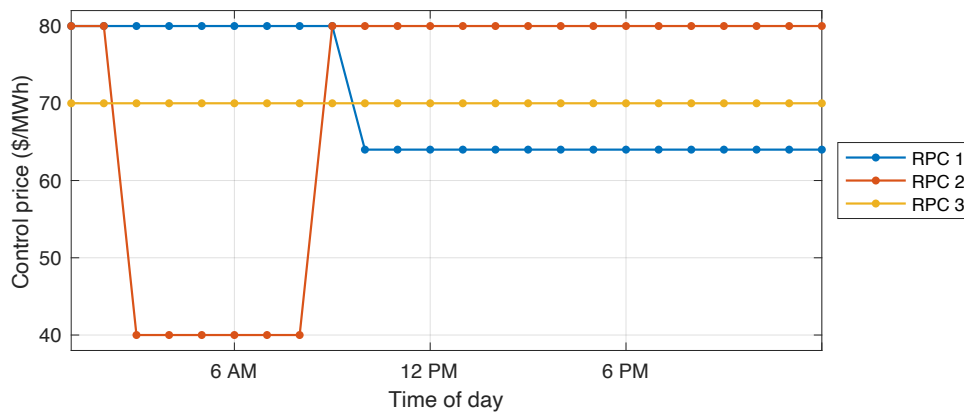

Figure 3.6. Control price signals to be broadcasted by the DSO according to the highest probability RPC of customers.

Figure 3.7 depicts the expected consumption profiles based on the corresponding price signals. Compared to the cluster center in Figure 3.4, customers with RPC 1 slightly increase their consumption in the afternoon. Instead, end-users using the RPC 3 show a more significant consumption during this time. However, cost minimization for them is only possible by shedding loads due to the uniform price. The same happens with the second group of customers, for which shifting their loads to the early morning is more difficult due to their inherent behavior.

Examination of the total net active and reactive power, shown in Figure 3.8, reveals that the DSO can control the customers' consumption and purchase higher power at lower market prices (the scheduled active power peak matches the lowest LMP at 5 PM). This finding confirms the effectiveness of the optimization model in scheduling DR. Interestingly, the reactive power exhibits some values that exceed the

corresponding active power values, influenced by the PV production and the higher consumption of low-power factor commercial end-users.
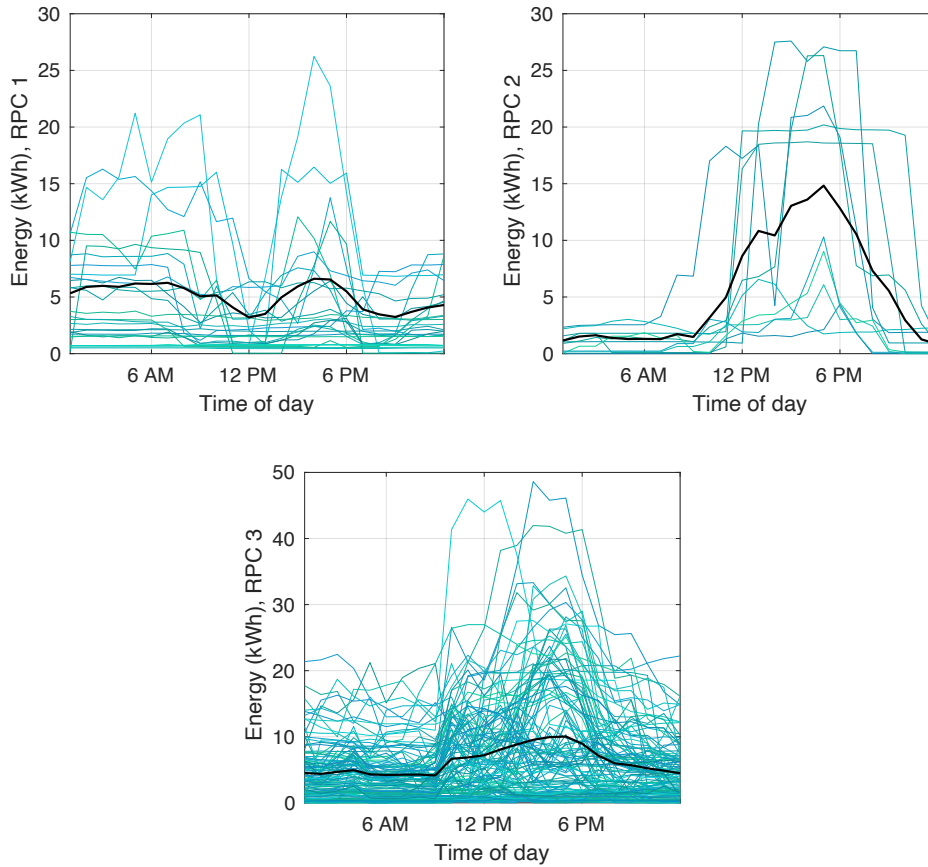

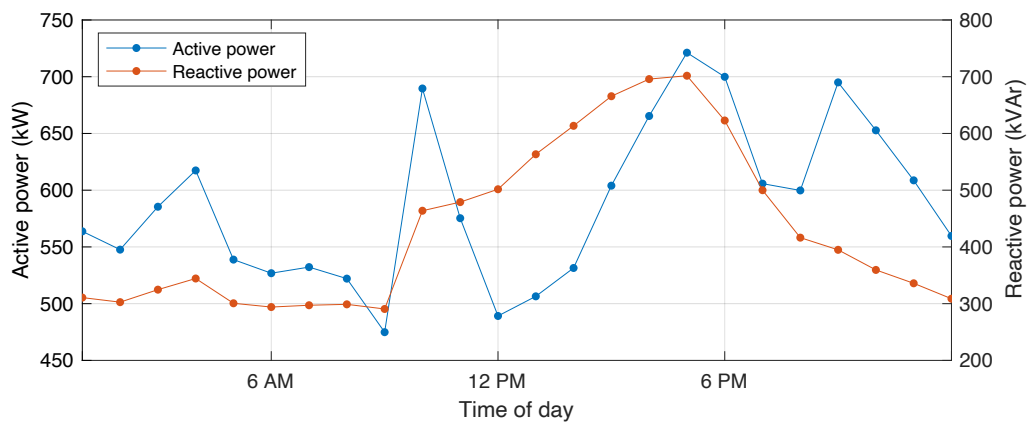Figure 3.7. Expected consumption profiles of customers, with the average in black.

Figure 3.8. Net active and reactive power demand from the electric power system.

Similarly, Figure 3.9 illustrates the expected voltage magnitudes for the distribution nodes. By inspection, only voltage drops happen during the day. As is typical in distribution systems, the farther nodes from the substation node 0 experience higher drops, which is the case, for example, of nodes 741, 740, and 711, from 10 AM to 2 PM. However, these voltage values are still above the minimum limit. In addition, voltage values from 3 PM are even lower, which is significant since uncertainty in load and generation is higher in this period.
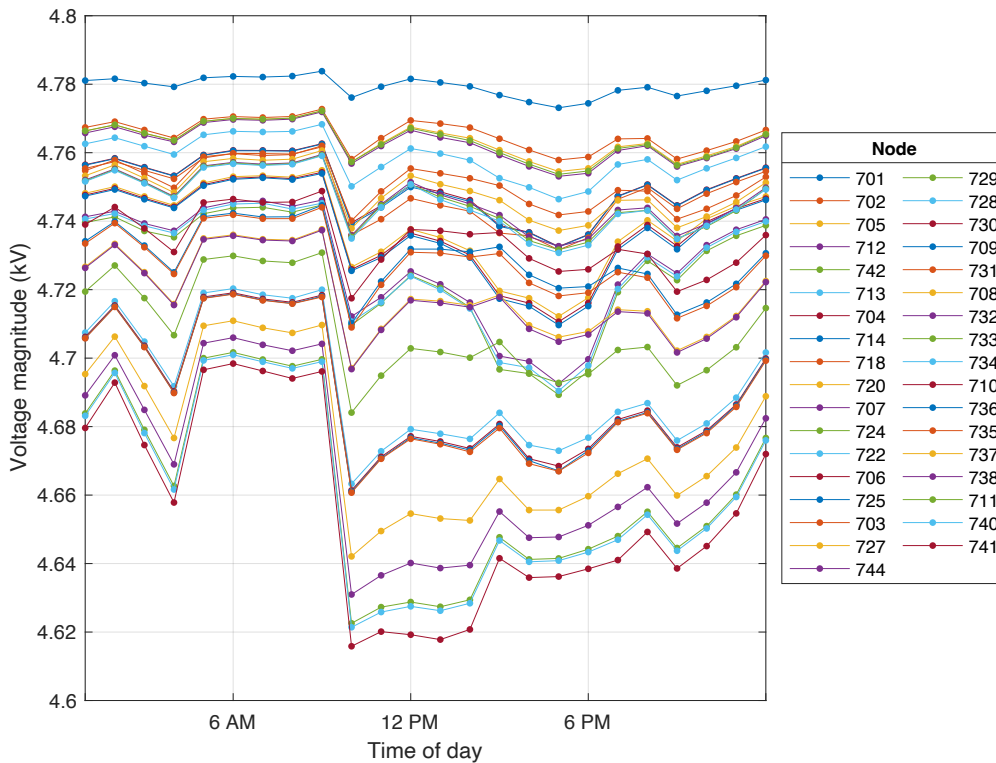
Figure 3.9. Voltage magnitudes in distribution nodes.

3.5.4 Analysis of the Uncertainty Cost

The uncertainty sources within the distribution system comprise the nodal-level load deviations and forecast errors of PV generation. In the following, this section examines the effect of uncertainty on nodal voltages.

As the influence of these power fluctuations differs throughout the system, Figure 3.10 investigates each nodal voltage by calculating the corresponding average uncertainty margin, which represents the decrease in available voltage caused by CCs, averaged

over the hours of the day. The higher uncertainty margins are necessary for nodes farther from the substation, associated with higher voltage drops. Specifically, nodes 724 and 722 are the most influenced by deviations and errors of load and generation.

Figure 3.10. Average uncertainty margins of voltage necessary to ensure the distribution system against uncertainty.
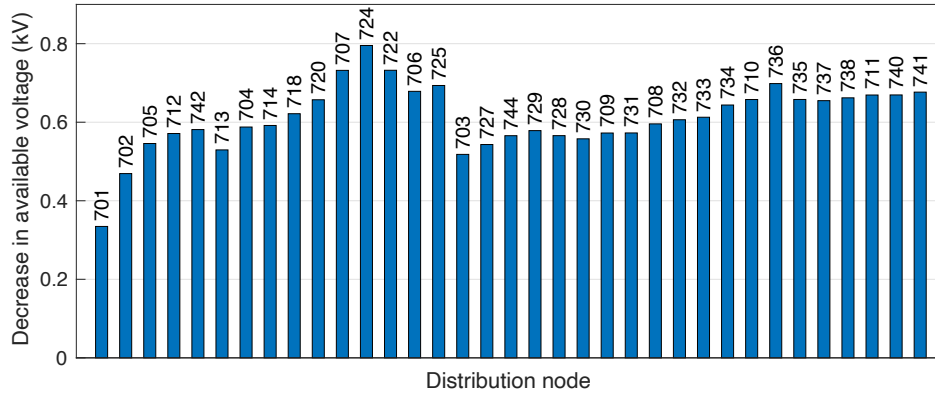
Additionally, Table 3.3 indicates the hours with the highest and lowest uncertainty margins. According to this outcome, power fluctuations influence especially from 3 PM, as expected, due to the higher uncertainty.

Nodes 740 and 741 are the two most critical in the distribution system due to their great distance from the substation. Figure 3.11 depicts two examples with these nodes that compare the resulting uncertainty margins with the empirical margins obtained from the distribution of voltage deviations considering time point $t = 17$ (5 PM), in which uncertainty has a high standard deviation and both nodes experience a significant voltage drop (as shown in Figure 3.9). As a result, the uncertainty margins in the study (based on the normal distribution), in red, match very closely to the empirical margins, in green. The most important outcome is that that difference be as minimal as possible, regardless of whether the normal distribution underestimates or overestimates the empirical margin.

Furthermore, the investigation presents a comparison analysis with the deterministic formulation to assess the approach's effectiveness concerning the profit for the DSO on the test day and the feasibility of the CCs in the solution. The deterministic formulation is applied, for instance, in [42]. For CCs, this formulation implies considering the nominal constraint, resulting in a mixed-integer linear programming (MILP) problem.

Table 3.3. Hours with the Highest and Lowest Uncertainty Margins of Voltage

Node	Hour of highest uncertainty margin	Hour of lowest uncertainty margin	Node	Hour of highest uncertainty margin	Hour of lowest uncertainty margin
701	6 PM	1 AM	729	7 PM	2 AM
702	6 PM	1 AM	728	11 PM	2 AM
705	6 PM	1 AM	730	3 PM	2 AM
712	3 PM	1 AM	709	3 PM	2 AM
742	6 PM	1 AM	731	3 PM	2 AM
713	11 PM	1 AM	708	10 PM	2 AM
704	11 PM	1 AM	732	10 PM	2 AM
714	6 PM	1 AM	733	10 PM	2 AM
718	6 PM	1 AM	734	10 PM	2 AM
720	5 PM	1 AM	710	3 PM	2 AM
707	11 PM	1 AM	736	3 PM	1 AM
724	11 PM	1 AM	735	3 PM	2 AM
722	11 PM	1 AM	737	10 PM	2 AM
706	5 PM	1 AM	738	10 PM	2 AM
725	5 PM	1 AM	711	10 PM	2 AM
703	10 PM	2 AM	740	10 PM	2 AM
727	10 PM	2 AM	741	10 PM	2 AM
744	11 PM	2 AM			

Considering the same critical nodes 740 and 741 for $t = 17$, Figure 3.12 shows the expected values of nodal voltage obtained with the MILP and MISOCP problems, with the cumulative distribution function (CDF) of voltage deviations superimposed. Due to the uncertainty margins in the MISOCP problem, the expected nodal voltage increases to $21.54 \, [\text{kV}]^2$ in the first case and $21.49 \, [\text{kV}]^2$ in the second case. This

increase represents a reduction of the probability of constraint violation from around 50 % in both cases with the MILP problem to less than the established 5 % with the proposed method. In particular, the expected value using the MILP problem for node 741 coincides with the minimum voltage limit.
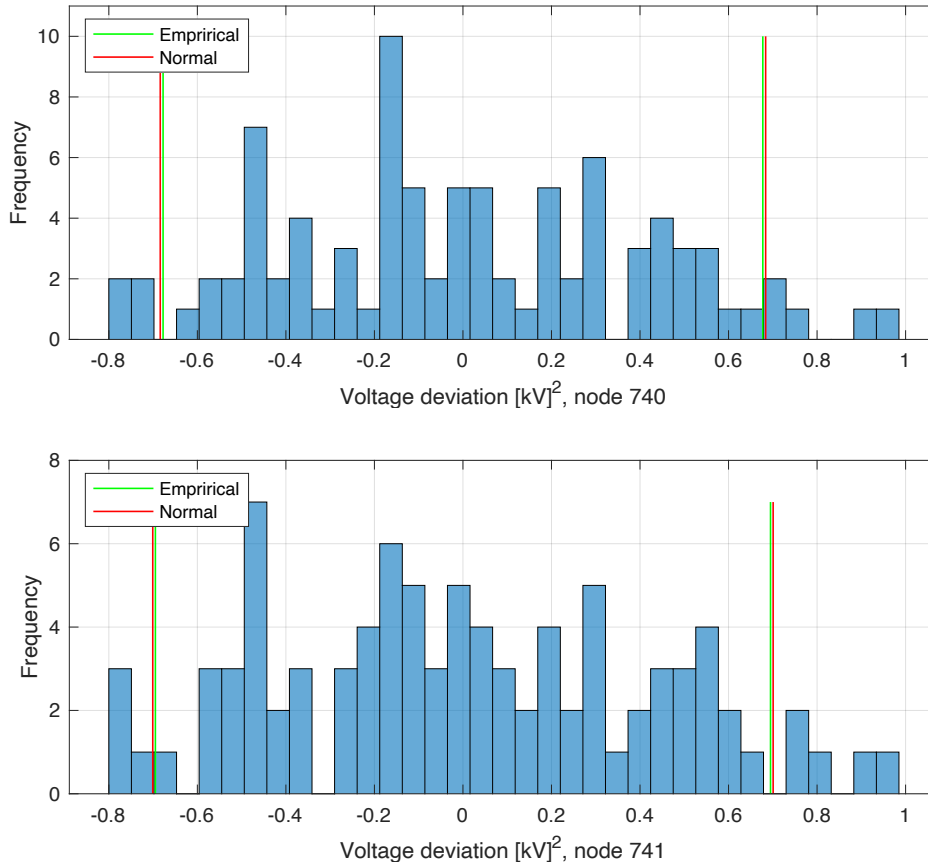
Figure 3.11. Histogram of voltage deviations for nodes 740 and 741 at 5 PM, where uncertainty margins are computed using the normal distribution and empirically.

On the other hand, Table 3.4 compares the DSO's expected profit with the deterministic case when different acceptable violation probabilities are employed. All chance-constrained solutions with different values for ϵ_v have a lower profit (or higher cost) than the MILP problem, which indicates that considering uncertainty increases the nominal cost of operation. However, the differences are negligible and tend to decrease with higher values of ϵ_v. In cases where ϵ_v is smaller than 0.05, the algorithm cannot find a feasible solution.
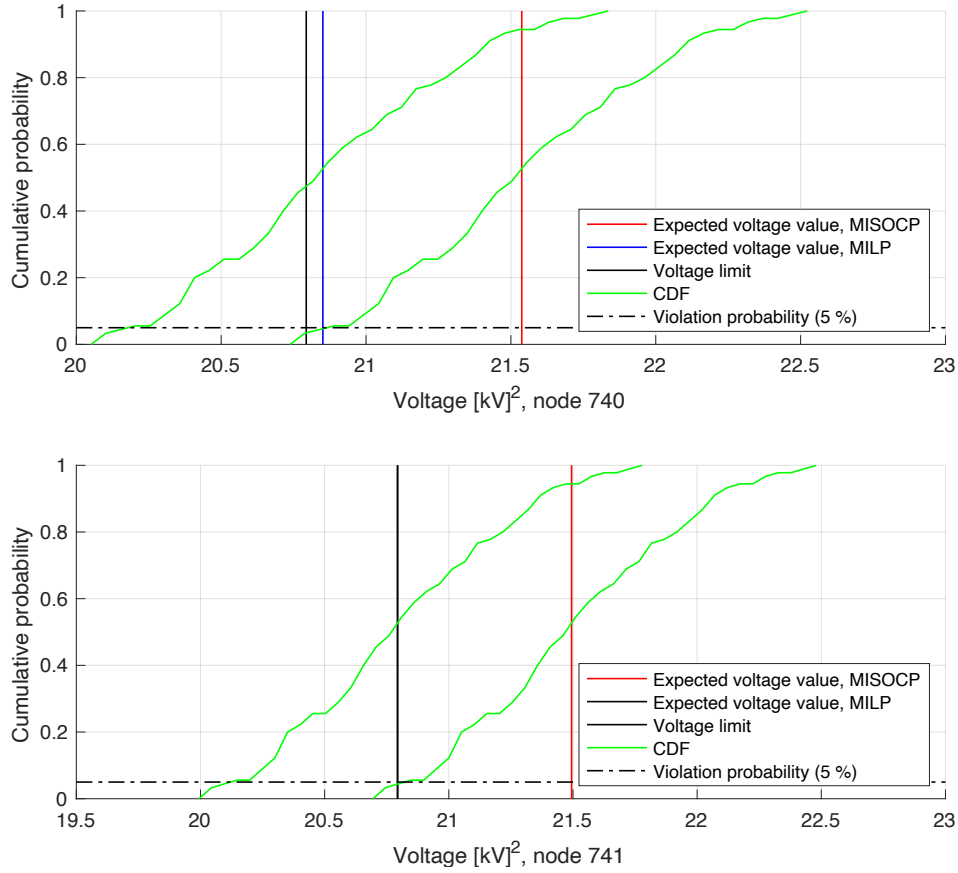
Figure 3.12. Expected voltage values on nodes 740 and 741 at 5 PM computed with the MILP and MISOCP problems, where the CDF of voltage deviations is superimposed on the expected values.

Table 3.4. DSO's Expected Profit with the MILP and MISCOP problems using Different Violation Probabilities of Voltage

	MILP	Violation probability $\epsilon_{\mathbf{v}}$			
		0.05	0.075	0.1	0.125
Profit ($)	579.4	576.1	577.1	577.5	577.8
Ratio regarding the deterministic	1	0.9943	0.996	0.9967	0.9972

Concerning the feasibility of CCs, Figure 3.13 illustrates the empirical probabilities of voltage violations for 720 new random samples of nodal load deviations and forecast errors obtained from a Monte Carlo simulation. The frequencies of voltage violations

with the MILP problem solution are much higher, where some constraints violate more than 50 % of cases, even with certainty in the available active power of PV facilities. This result highlights the need for an uncertainty-aware approach. All chance-constrained solutions have lower violation probabilities. Although there are always some constraints with an empirical probability of violation higher than the acceptable value, the difference is minor.
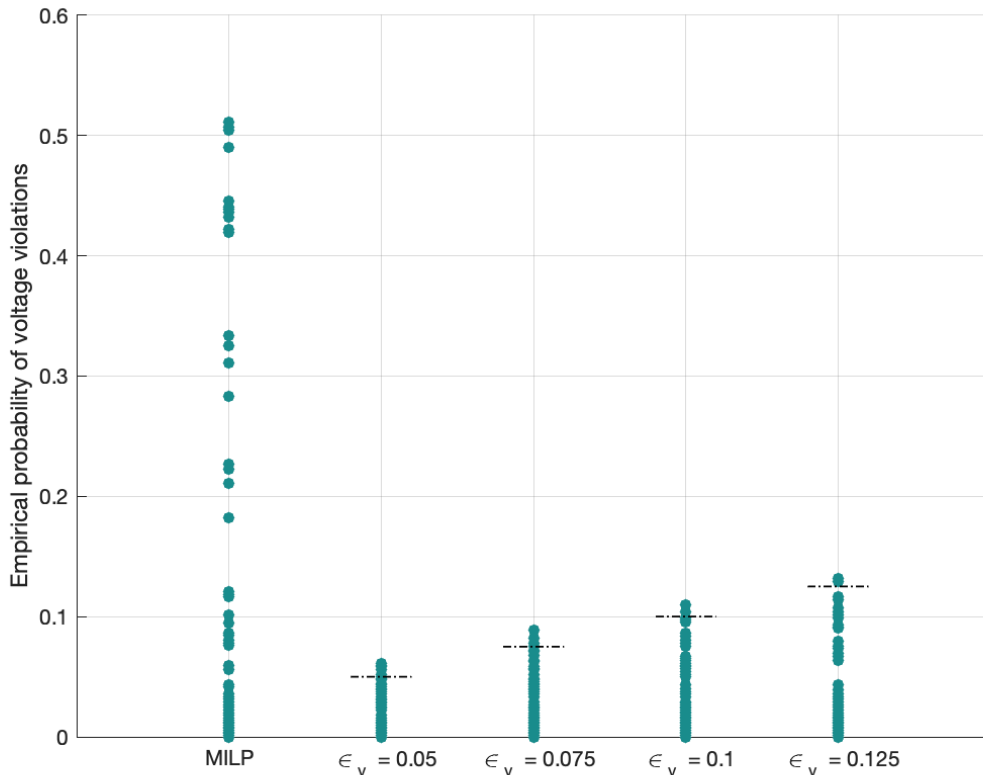
Figure 3.13. Empirical probabilities of voltage violations for 720 new random samples of nodal active power deviations and errors on each hour with the MILP and MISOCP problems, using different values of ϵ_v.

Finally, contrary to the voltage case, violations of power flows are only a few and below the adopted limit due to the loose power limitations that characterize the distribution lines of the test feeder. However, to show the effect of CCs, Figure 3.14 depicts the maximum uncertainty margins for each distribution line considering all time points. As a result, larger values are necessary for lines at the beginning of the grid, as they

are influenced directly by load deviations and forecast errors in the downstream nodes they supply.

Figure 3.14. Maximum uncertainty margins of lines (identified by the receiving nodes) necessary to ensure the distribution system against uncertainty.

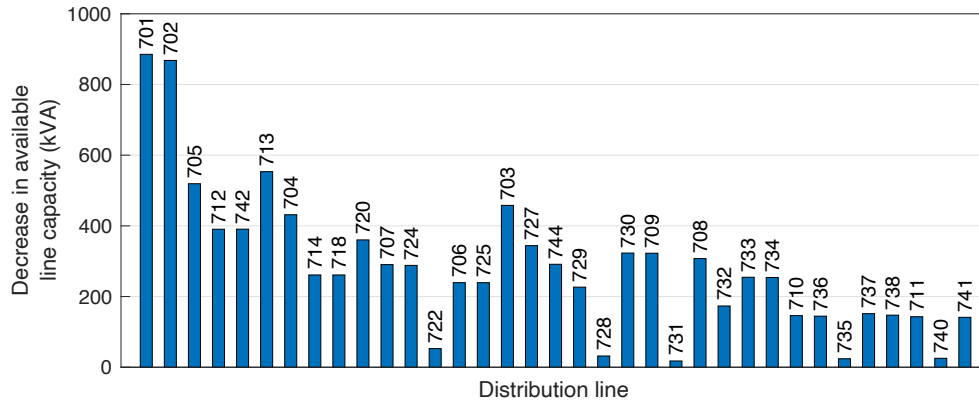## 3.6 Conclusion

The main idea of this chapter is to propose an online framework for scheduling customers' DR to support the integration of PV generation into the distribution system. In particular, the DSO can use the proposed approach to obtain financial profits and maintain the reliable operation of the distribution grid. The approach uses a chance-constrained bi-level optimization model that delivers control price signals for different groups of customers. Furthermore, CCs enforce voltage and power flow limits with a predefined violation probability, and the uncertainty margin in their reformulations mitigates the effect of power deviations.

The conducted case study demonstrates the suitability of the approach for energy management on the demand side under uncertainty. According to the three RPCs of consumers for the test day, the algorithm provides the corresponding control price signals. Concerning the uncertainty cost, results highlight that by using adequate violation probabilities, the DSO ensures the reliable operation of the distribution system with a profit value almost identical to the deterministic. Finally, although the normal distribution does not guarantee that the empirical probabilities of voltage violations are less than the violation probability, the differences are minor. Therefore, the approach can be applied in practical situations when the violation probability does not represent a hard limit.

Chapter 4

Online Demand Response and Photovoltaic Inverter Dispatch Scheduling Considering Uncertainty

This chapter expands the previous online framework for the customized scheduling of customers' demand response (DR) alongside the power dispatch of distribution-level photovoltaic (PV) generation facilities. The proposed model enables these energy resources to respond to the service required by the distribution system operator (DSO) by determining optimal control price signals for groups of customers and active and reactive power set points for PV inverters, which ensures the reliable operation of the distribution grid while maximizing the DSO's profit. The approach also addresses uncertainty in the distribution system modeling by incorporating, in addition to the above, chance constraints (CCs) for the apparent power of PV inverters, for which the analytical reformulation derives computationally feasible convex constraints. Finally, the chapter presents a case study for scheduling residential and commercial Chilean end-users and PV facilities using the IEEE-37 node test distribution feeder and real-world local market prices and daily profiles of customers.

4.1 Motivation

High penetration of PV generation can introduce voltage regulation problems in distribution systems, and traditional volt-var devices can be insufficient to adapt to the dynamics of PV generation to offer the desired voltage performance. An alternative solution takes advantage of the inverter's control capability. The primary function of PV inverters is to maximize active power generation. However, since this active power output is time-varying, PV inverters do not always reach their capacity limit. Therefore, they can also be employed to either absorb or inject reactive power, which helps to regulate the local voltage [72].

Recently, the implementation of optimization approaches that include computing the power set points for PV inverters under uncertainty has been developed based on forecasts of available generation and expected customers' behavior. Important related works involve, for instance, [54], which presents an adaptive chance-constrained AC OPF formulation where the set points for the PV and battery systems are optimized while enforcing voltage regulation with uncertainty in both RESs and loads. Similarly, [73] allows the DSO to obtain an optimal dispatch plan for PV and battery systems using a distributionally robust chance-constrained model with an ambiguity set for uncertainties of PV production capability, end-user consumption, requested flexibility by the external grid operator, and squared voltage magnitude at the point of common coupling. Furthermore, [74] formulates a two-stage robust optimization model that, considering the uncertainty of PV generation, selects the critical subset of PV inverters to provide ancillary services to the grid and finds their optimal dispatching real and reactive power set points. Finally, accounting for uncertainties in demand and PV generation, [75] develops a reactive power control method that performs two tasks: 1) reducing the active power loss of the system by periodically dispatching the inverter's reactive power set points, and 2) solving the overvoltage problem by applying a real-time volt-var algorithm.

The above studies focus on determining the optimal inverter's power set points, mainly for local voltage regulation, while considering the uncertainty in PV production. Although most of them ([54], [73], and [75]) also include uncertainty in customer behavior, these investigations do not intend to control DR.

This chapter expands the previous online framework proposed in Chapter 2 by exploiting the reactive power control the PV inverters can provide in the distribution grid. Thus, aiming at the DSO's profit maximization, this framework schedules both customers' DR through the dynamic price signals and PV facilities based on an optimal dispatch of inverters' active and reactive power set points. The proposed algorithm also accounts for the uncertainty in the distribution-system modeling, including that related to the operation of PV facilities, for which the work uses the presented analytical reformulation (see Appendix B) to obtain computationally feasible convex constraints. Finally, the proposed model is tested in a case study using the IEEE-37 node test distribution feeder with local market prices and daily profiles of the residential and commercial Chilean end-users.

The main contributions of the chapter are summarized below:

1) An extension of the chance-constrained bi-level model to include determining optimal power set points for dispatchable inverters of PV facilities is proposed such

that squared nodal voltage, complex power flow in lines, and inverter capacity limits are satisfied with high probability.

2) The framework is tested in a case study with the real-world market prices and daily profiles of Chilean residential and commercial end-users on the IEEE-37 node test distribution feeder. Results demonstrate how the presented framework enables optimal scheduling of flexible customers and PV facilities while considering both the previous uncertainty and uncertainty concerning the operation of PV facilities.

The organization of this chapter is as follows: Section 4.2 defines the capability of the PV facility, which limits the injection or absorption of the inverter's reactive power. Section 4.3 introduces the CC for PV production and develops the corresponding reformulation. The case study using the IEEE-37 node test distribution feeder is presented in Section 4.4 to test the proposed framework and analyze the uncertainty cost. Finally, Section 4.5 concludes the chapter.

4.2   Capability Constraints for the PV Facility

With the increasing penetration of distribution-level PV facilities, interest is moving toward using inverters' capability to absorb or inject reactive power as needed. In addition to (3.1) that allows a flexible active power value for the PV production facility g at time point t, the following operational constraint defines the inverter's capability for also adjusting the reactive power output:

$$\bar{p}_{g_t}^{\ 2} + \bar{q}_{g_t}^{\ 2} \leq S_g^{\max 2}, \quad g \in \mathbf{G}, t \in \mathbf{T}, \qquad (4.1)$$

where $S_g^{\max}$ denotes the inverter's rated apparent power of PV facility g.

However, the recently emerged IEEE Standard 1547 [76], which encourages the inverter-level modulation of power values in response to local grid conditions, recommends injecting or absorbing reactive power for active power output levels greater than or equal to the minimum steady-state active power capability. Therefore, the convex region defined from this value in the complex plane entitles the reactive power production exclusively during daylight.

4.3   Reformulation of CCs for the PV Facility

Due to the stochastic nature of PV generation, the thesis formulates the following set of constraints for inverters' power set points to be incorporated into the chance-constrained bi-level model of Section 3.2.3:

$$\mathbb{P}\left\{p_{g_t}^2 + q_{g_t}^2 \leq S_g^{\max 2}\right\} \geq 1 - \epsilon_{\mathrm{g}}, \quad g \in \mathbf{G}, t \in \mathbf{T}, \quad (4.2)$$

where q_{g_t} is the reactive power of PV facility g at time point t and ϵ_{g} the corresponding violation probability.

Therefore, given the response profile classes (RPCs) of customers, the forecast of the available active power in PV facilities and the locational marginal price (LMP), CCs in (3.15), (3.16), and now (4.2) ensure that scheduling customers' DR and inverters' power set points satisfies the squared nodal voltage, complex power flow in lines, and inverter capacity limits with high probability (in this last case with probability $1 - \epsilon_{\mathrm{g}}$).

4.3.1 Reformulation of the CCs

Every single constraint in (4.2) with a limit for the inverter capacity is a CC of the general form:

$$\mathbb{P}\left\{\left[a_1(\lambda_{l_t}) + \xi_{g_t}\right]^2 + \left[a_2(\lambda_{l_t})\right]^2 \leq c^2\right\} \geq 1 - \epsilon, \quad (4.3)$$

where the squared random variable ξ_1 and the squared scalar variable ξ_2 on the left-hand side represent the operating points of the active and reactive power of PV production contained in the feasible region $\mathcal{S} \in \mathbb{R}^2$, while c is a constant representing the inverter's apparent power limit.

Since $\mathcal{S}$ is convex, it can be inner approximated by the I-sided polygon $\mathcal{P}$ inscribed inside $\mathcal{S}$, as in the case of CCs for active and reactive power flows in Section 3.4. By expanding on this, the following linear CC is generalized:

$$\mathbb{P}\left\{w_{i,1}\left[a_1(\lambda_{l_t}) + \xi_{g_t}\right] + w_{i,2}a_2(\lambda_{l_t}) + w_{i,3}c \leq 0\right\} \geq 1 - \epsilon, \quad \forall i: w_{i,1} \geq 0 \quad (4.4)$$

CCs in (4.4) can finally be converted into one with the same structure as (3.18) with the random variable ξ_{g_t} of normal distribution. Thus, the following set of constraints for the half-space approximation of $\mathcal{S}$ can replace the CCs of inverters' active and reactive power set points in (4.2):

$$w_{i,1}\bar{p}_{g_t} + w_{i,2}\bar{q}_{g_t} \leq -w_{i,3}S_g^{\max} - \Phi^{-1}\left(1 - \epsilon_g\right)\left\|w_{i,1}\sigma_{g_t}\right\|_2, \quad \forall i: w_{i,1} \geq 0, g \in \mathbf{G}, t$$
$$\in \mathbf{T}, \tag{4.5}$$

where σ_{g_t} is the standard deviation of forecast errors for PV facility g at time point t.

4.4   Case Study

This section presents a case study that expands the previous analysis in Section 3.5 to include the power dispatch of distribution-level PV generation facilities. First, the section performs a characterization stage using electricity data from SMs of 125 residential and commercial Chilean end-users to determine the RPCs and their combinations. Second, an application example provides optimal control prices for customers and active and reactive power set points for PV inverters. Finally, the section analyzes the impact of CCs concerning inverter capacity limits in the optimization problem.

The electricity data corresponds to the three weeks of February 14 to March 5, 2020; therefore, the test day uses the original LMP of March 6, 2020, associated with the substation of customers. Using the IEEE 37-node test distribution feeder, Table 4.1 summarizes the allocation of consumers in load nodes and the nodes for the same two solar facilities, which in this chapter are located farther from the substation. For simplicity, in both inverters of PV facilities, $S_g^{\max} = 350$ kVA, whereas the minimum steady-state active power capability considers a value of 7 % of $S_g^{\max}$. Lastly, the violation probabilities in CCs are $\epsilon_v = 0.1$, $\epsilon_i = 0.01$, and $\epsilon_g = 0.01$, while the rest of the parameters and variables remain the same as in Section 3.5.

4.4.1   Customer Characterization and Modeling

The application of the CFSFDP to the set of daily profiles allows the identification of three clusters through the corresponding plots. Figure 4.1 illustrates the result and highlights the selected centers with colored and bigger dots. One (with the violet color) is remarkably different because of its higher neighborhood compared to the other two.

Figure 4.2 shows the corresponding daily profiles of clusters, with cardinalities of 1348, 275, and 993, respectively. Cluster 1 involves profiles with a typical residential pattern of slightly higher consumption at the end of the day. Cluster 2 presents a pattern of lower consumption during daylight hours due to probably empty households on those (vacation) days. Conversely, cluster 3 exhibits a higher consumption in those hours,

with two notable peaks around 12 PM and 6 PM. Specifically, cluster centers 1, 2, and 3 (in black) correspond to the violet, brown, and green dots in Figure 4.1.

Table 4.1. Number of Customers and PV Facilities in Nodes of the Test Feeder

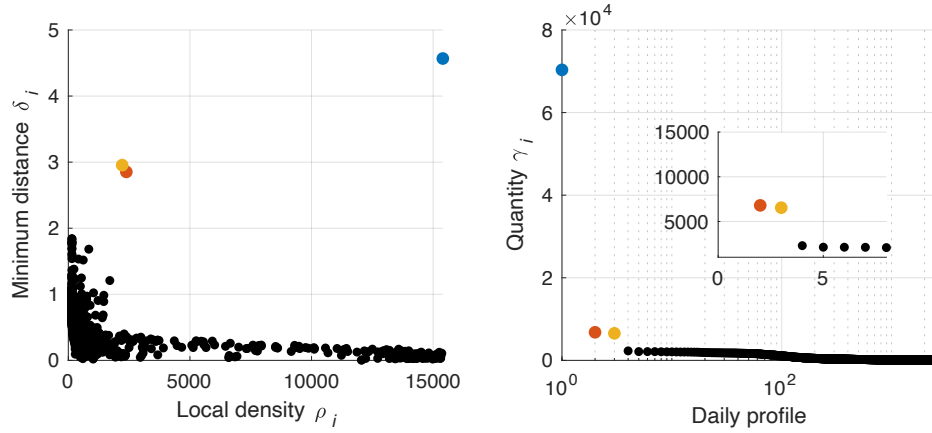| Load node	Number of customers	Generation node	Number of PV facilities
701	38	707	1
712	6	711	1
713	6		
714	4		
718	6		
720	3		
722	6		
724	1		
728	3		
730	3		
731	2		
732	1		
733	3		
734	1		
735	3		
736	1		
737	4		
738	4		
740	2		
741	22		
742	6		

Figure 4.1. Plots for identification of clusters: the decision graph and quantities γ in decreasing order (including a zoom to better distinguish the last two).
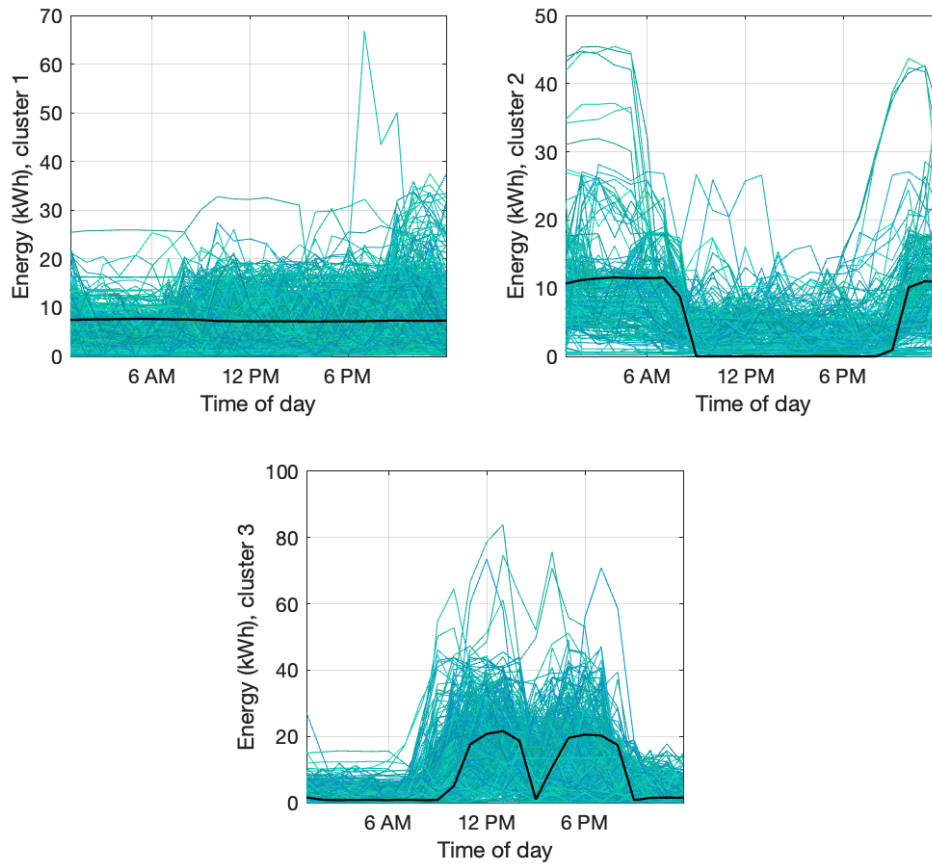

Figure 4.2. Daily consumption profiles of clusters, with cluster centers in black.

Table 4.2 summarizes the RPCs and the combinations of these daily patterns. Again, most customers use two RPCs during the period, mainly consumption patterns 1 and 3.

Table 4.2. RPCs of Customers and their Combinations

RPC	1	2	3
Number of customers	21	76	28
Combination of RPCs (Number of customers of the combination)	1 (14) 2 (3) 3 (4)	1-2 (15) 1-3 (59) 2-3 (2)	1-2-3 (28)

As in the previous analysis, it is assumed in this chapter that consumers use the RPC with the highest probability for the test day, which results in 64 customers using the RPC 1, 7 the RPC 2, and 54 the RPC 3. Figure 4.3 depicts the customers' aggregate average demand from their corresponding RPC, the available PV generation profile, and the market prices for the test day.
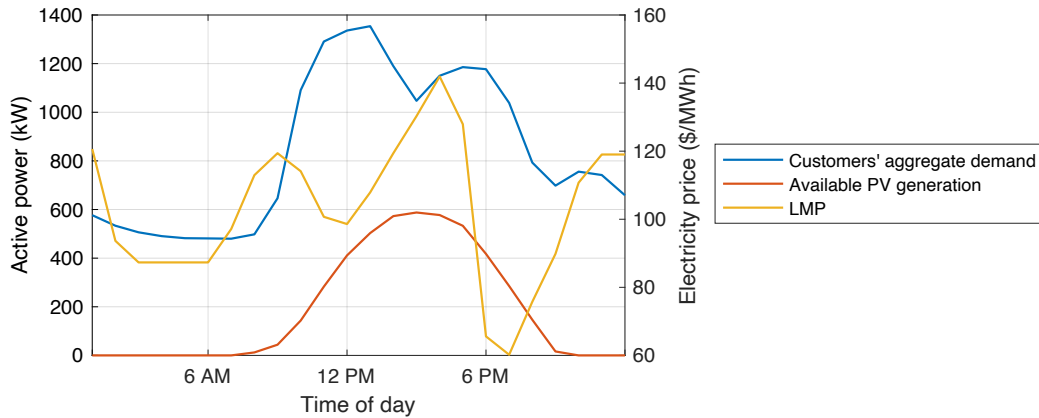

Figure 4.3. Expected customers' aggregate average demand, available PV generation, and LMP on March 6, 2020.

4.4.2 DR Scheduling with Optimal Dispatch of PV Inverters

Application of **Algorithm 2** can include determining power set points for PV inverters. Considering time point $t = 1$, Figures 4.4 and 4.5 illustrate the three control price signals to be broadcasted by the DSO and the expected consumption profiles of end-users, respectively. For customers using RPC 1, the maximum price at 10 PM and 11 PM encourages them to bring forward their intended activities. As Figure 4.5 verifies, it is easier for this group to shift some activities to a closer period (one to three hours before 10 PM), which coincides with the lower market prices. With **80 $/MWh** at the beginning and the end of the day, the second control price signal seeks that consumers with the second pattern carry out their activities mostly between 9 AM and 9 PM, where they consume less, but the PV generation is high. Compared to the other two groups, these customers have a low influence since they represent only 5.6 % of the total. Finally, the last control signal aims to flatten the double-peak consumption profiles of end-users with RPC 3 through a price of approximately **66 $/MWh** from 5 PM and before 9 AM, which involves the hours with the lower market prices. According to the shapes in Figure 4.5 for this third group, their inherent flexibility allows these customers to have a peak consumption around 7 PM, coinciding with the lowest market price and thus contributing to a high expected profit for the DSO.
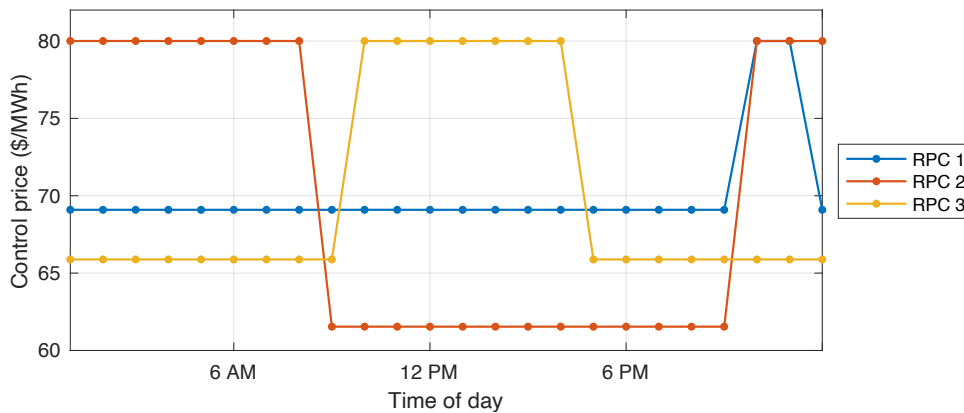
Figure 4.4. Control price signals to be broadcasted by the DSO according to the highest probability RPC of customers.

Figure 4.6 depicts the active and reactive power set points of PV inverters. With the connection of the PV facilities at the farther nodes 707 and 711, the active power output produces some values below the available active power during the higher generation period because of the uncertainty margin considered. However, as Figure

4.7 shows, voltage magnitudes in these nodes are generally closer to the nominal value of 4.8 kV (the same for their corresponding downstream nodes), even exceeding it in hours of high generation. For the reactive power, based on the recommendation given by IEEE Standard 1547, injections occur mainly during hours of lower active power production.
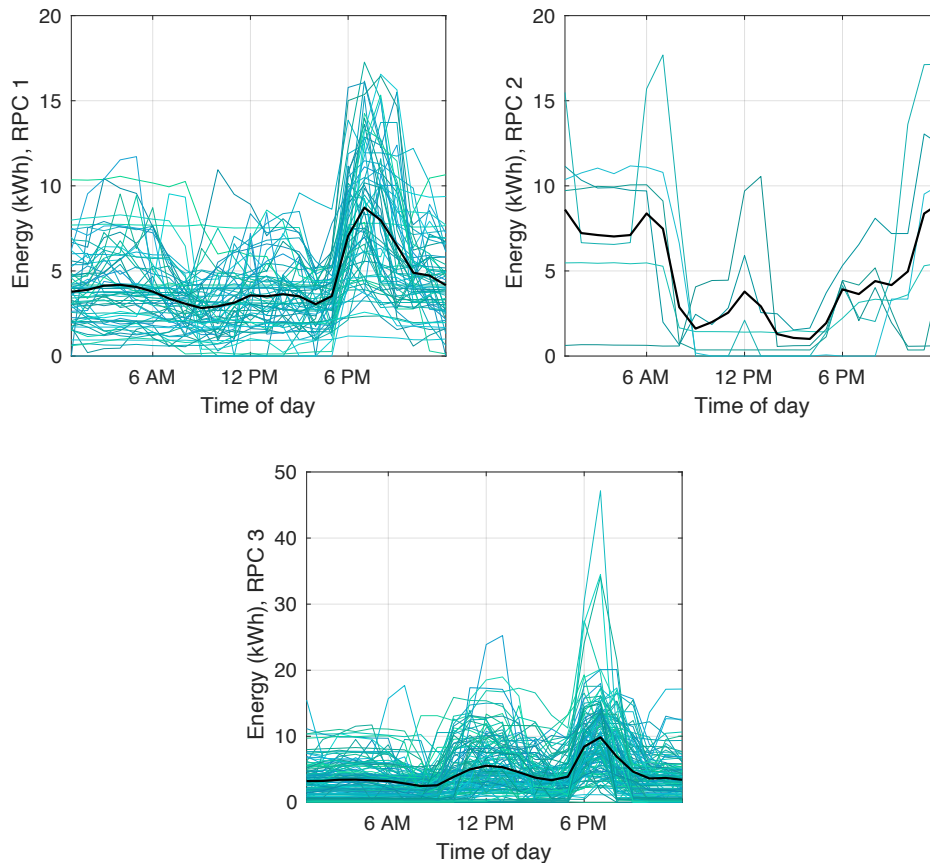
Figure 4.5. Expected consumption profiles of customers, with the average in black.

Figure 4.8 illustrates the maximum uncertainty margins and a comparison with the mixed-integer linear programming (MILP) formulation of the empirical probabilities of inverters' apparent power violations. For the empirical probabilities, the case study uses 720 new random samples of forecast errors each hour obtained from a Monte Carlo simulation. Based on the uncertainty margin considered, the decrease in available apparent power reaches the presented highest values at 3 PM and 4 PM for nodes 707 and 711, respectively. As expected, the deterministic solution shows many apparent power violations that exceed the acceptable limit. Like the case with nodal

voltages in Chapter 3, some constraints have slightly higher empirical probabilities, but the difference is minor.
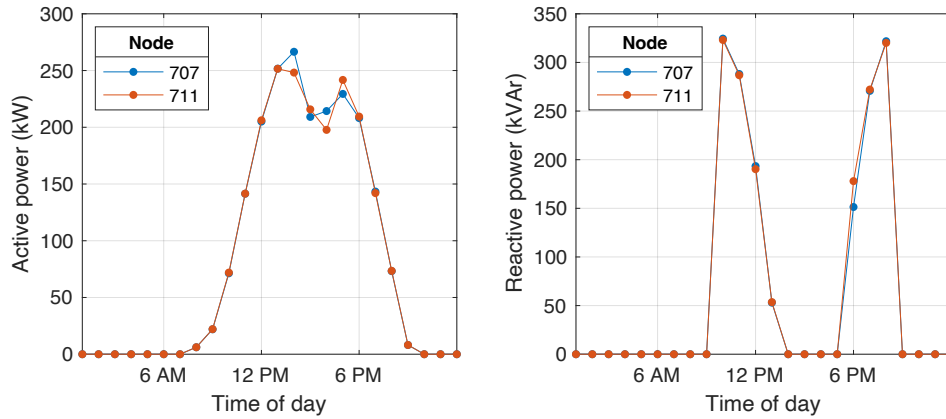
Figure 4.6. Active and reactive power set points for inverters of PV facilities at nodes 707 and 711.
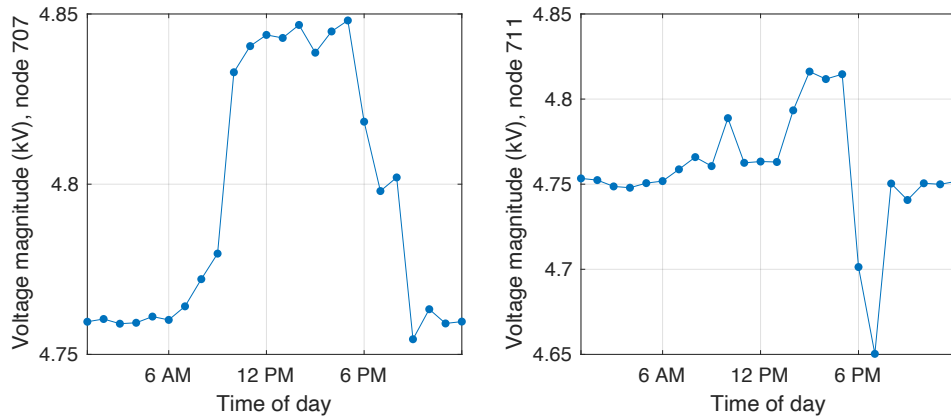
Figure 4.7. Voltage magnitudes in distribution nodes 707 and 711.

Finally, Table 4.3 summarizes the DSO's expected profit for both situations, from which a remarkable difference between the two solutions results compared with the cases analyzed in Table 3.4. The lower profit in this case study is due to the inclusion of the uncertainty margin associated with the operation of PV facilities.

4.5   Conclusion

The main idea of this chapter is to expand the proposed framework for scheduling customers' power responses to include the optimal dispatch of PV generation facilities.

Therefore, aiming at a profit maximization objective and ensuring a reliable operation of the distribution grid under uncertainty, the DSO determines optimal control prices for groups of customers and active and reactive power set points for PV inverters. In addition to CCs for voltages and power flows, the presented approach introduces and reformulates CCs for inverters' capacity.
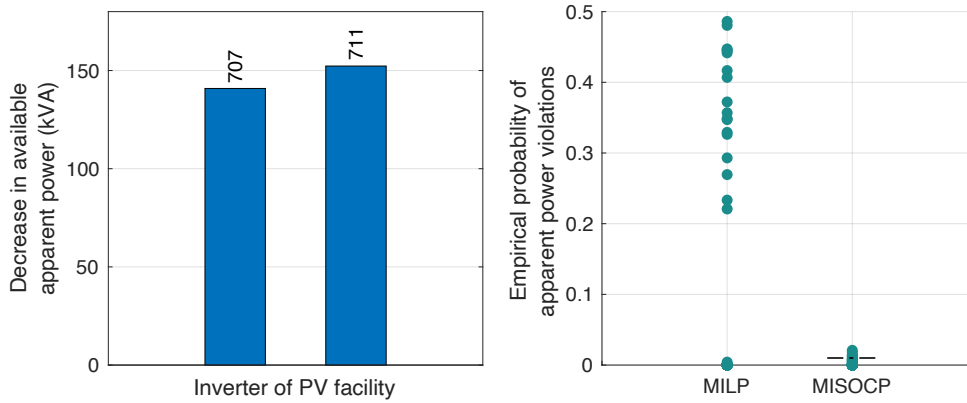
Figure 4.8. Maximum uncertainty margins of inverters and empirical probabilities of apparent power violations (including both PV facilities) for 720 new random samples of active power errors each hour with the MILP and MISOCP problems.

Table 4.3. DSO's Expected Profit with the MILP and MISCOP problems

	MILP	MISCOP
Profit ($)	152.9	93.4
Ratio regarding the deterministic	1	0.6106

Results of the conducted case study highlight the suitability of the approach for management distribution-level energy resources by providing the optimal price signals considering the three RPCs of consumers for the test day and the optimal use of the PV inverters' capability during daylight. Finally, apparent power violations close to the acceptable value ensure the secure scheduling of the distribution system exploiting inverters' active and reactive power set points, although this results in a lower profit for the DSO.

Chapter 5

Conclusion and Outlook

5.1 Summary

This thesis proposes a methodology for demand response (DR) scheduling based on the customized online learning of customers' behavior and investigates its application in the distribution system operation with uncertainty.

In Chapter 2, the thesis proposes an online framework for DR characterization over time. For the distribution system operator (DSO), the approach allows obtaining and updating the response profile classes (RPCs) and the variability of customers, but also estimating the customer response to the price signal based on a known RPC.

In Chapter 3, the thesis proposes an online framework for scheduling customers' power responses to support integrating photovoltaic (PV) generation into the distribution system. The approach is essential for the DSO by providing customized price signals for DR management. The DSO can also ensure the reliable operation of the distribution grid by dealing with uncertainties in demand behavior and PV production using chance constraints (CCs) for squared nodal voltage and complex power flow.

In Chapter 4, the thesis expands the previous online framework to include the optimal dispatch of PV facilities. Therefore, the DSO can determine control prices for groups of customers and active and reactive power set points for PV inverters while ensuring the reliable operation of the distribution grid by including, in addition to the above uncertainties, the corresponding to the operation of PV inverters.

5.2 Conclusions

The most important conclusions in the thesis are summarized below:

1) The application of an incremental methodology for the continuous processing of daily load profiles of customers can help to characterize their behavior over time. The presented online framework in Chapter 2 allows the DSO to obtain and update online mathematical models of customers. Therefore, the DSO can estimate the customers' power response to price signals according to demand preferences and environmental factors.

2) Customers' individual preferences and environmental factors influence their energy demand, which produces deviations from expected behaviors. However, the approach in Chapter 2 contributes to modeling the underlying probability distribution of these random deviations since each daily deviation can be considered a realization of the corresponding random variable. Using the parameters of this empirical distribution overcomes the limitation of making distributional assumptions that can result in risky or more conservative costly solutions. Investigation into model stochasticity in demand is scarce in the technical literature.

3) Traditionally, in both practice and technical literature, investigation of the active role of demand has little consideration. Under the concept of DR, however, the DSO can exploit the inherent flexibility from the demand side and contribute to integrating renewable production into the distribution grid for a more sustainable operation. The presented approach in Chapter 3 aims to that point by providing the (indirect) control of customers' power responses through optimal time-varying price signals according to their RPCs and considering the operating conditions of the distribution grid. Thus, an optimal solution is delivered, founded on the customized analysis of customers.

4) Generally, reformulating a chance-constrained programming problem into a tractable one is challenging. In most cases, guaranteeing both a feasible and optimal solution is unattainable. This thesis presents a methodology with a theoretical foundation but includes accepted assumptions necessary in practice. For example, the work uses single CCs and assumes the normal probability distribution, supported since the stochastic problem comprises many random variables with many realizations. The selection of the risk level, measured in CCs in terms of the violation probability, is also problematic. Higher security requirements (which imply more conservative violation probabilities) inherently increase the operational costs (or decrease the DSO's profit in the context of the thesis). An acceptable level of risk is thus not a fixed value but depends on the tradeoff between profit and risk exposure. Chapters 3 and 4 introduce CCs with the corresponding analytical reformulation for nodal voltages, power flows in lines, and PV inverters' capacity and analyze their interpretation in results.

5) With the increasing penetration of distribution-level PV facilities, interest is moving toward using inverters' capability for absorbing or injecting reactive power. Optimal dispatch of active and reactive power can contribute to local voltage regulation. Chapter 4 presents a framework that schedules customers' power responses through customized control price signals and PV inverters' power set points while ensuring a reliable operation of the distribution grid by dealing with uncertainties in nodal voltages, power flows in lines, and available active powers of PV facilities. The approach provides an expected profit for the DSO; however, the DSO has the final decision concerning the tradeoff between the distribution system's security and profit.

5.3 Outlook

The thesis's results provide the following directions for future work:

1) For online DR characterization, the impact of dimensionality on clustering needs to be explored in more depth. Furthermore, a suitable method to infer the daily RPC for estimating the expected response of customers is required, which is not necessarily the RPC with the highest probability.

2) This thesis focuses on a balanced distribution system for online DR scheduling. However, the practical background of the investigation addresses unbalanced load with distributed PV integration into the distribution grid, where violations of voltage, power flow, and inverter capacity limits can be present. Therefore, future work requires analysis of large-scale unbalanced distribution grids with the corresponding challenge in convex approximations for CCs.

3) It is significant to highlight the advantages of the online scheduling approach regarding balancing solution efficiency and global optimization. Using the model predictive control for online scheduling also requires further examination.

Appendix A

Calculation of Incremental Validity Indices

This study implements the iDB and iXB validity indices to monitor the performance of the online algorithm. First, the iDB index [32] results as follows:

$$
\text{iDB} = \frac{1}{K} \sum_{k=1}^{K} \max_{\mathbf{G}_m} \frac{1}{d_2(\mathbf{c}_k, \mathbf{c}_m)} \left(\frac{1}{N_k} \sum_{i=1}^{N_k} \left\| \mathbf{p}_{k,i} - \mathbf{c}_k \right\|_2 + \frac{1}{N_m} \sum_{j=1}^{N_m} \left\| \mathbf{p}_{m,j} - \mathbf{c}_m \right\|_2 \right), \quad k, m
$$
$$
= 1, \dots, K, \tag{A.1}
$$

which means the average of the similarity measures of each cluster k with its most similar one. The d_2 metric between the cluster centers indicates separation, whereas the expression in parenthesis shows their dispersions (compactness measure).

A hard partition version of the iXB index [32] that defines the ratio of compactness to separation is as follows:

$$
\text{iXB} = \frac{\sum_{k=1}^{K} \sum_{i=1}^{N_k} \left\| \mathbf{p}_{k,i} - \mathbf{c}_k \right\|_2^2}{N \min_{\mathbf{c}_m, \mathbf{c}_o} \left\| \mathbf{c}_m - \mathbf{c}_o \right\|_2^2}, \quad k, m, o = 1, \dots, K, \tag{A.2}
$$

where N is the total number of daily profiles.

For each cluster k that does not undergo splitting or merging, the new compactness is defined as:

$$\varsigma_k = \sum_{i=1}^{N_k^{\mathrm{p}}} \left\| \mathbf{p}_{k,i} - \mathbf{c}_k \right\|_2^q + \sum_{j=1}^{N_k^{\mathrm{c}}} \left\| \mathbf{p}_{k,j} - \mathbf{c}_k \right\|_2^q, \quad 1 \leq k \leq K, \qquad (\mathrm{A}.3)$$

where $q = 1$ for the iDB index and $q = 2$ for the iXB index.

Equation $(\mathrm{A}.3)$ separates the compactness value into the objects up to the previous day N_k^{p} and the effect of the new objects N_k^{c} added to the cluster k. Therefore, from ς_k, it is possible to obtain both indices by applying $(\mathrm{A}.1)$ and $(\mathrm{A}.2)$, respectively.

Appendix B

Analytical Reformulation of Chance Constraints

For the solution to the chance-constrained bi-level optimization model in (3.12), CCs in (3.15), (3.16), and (4.2) for magnitudes of squared nodal voltages, complex power flows in lines and inverters' capacity require reformulating into deterministic and tractable constraints. These constraints are single CCs of the general form expressed in (3.18).

Constraint (3.18) can be represented equivalently as:

$$\mathbb{P}\left\{\frac{\xi - \mu_\xi(\lambda_{l_t})}{\sigma_\xi(\lambda_{l_t})} < \frac{c - \mu_\xi(\lambda_{l_t})}{\sigma_\xi(\lambda_{l_t})}\right\} = \mathbb{P}\left\{\xi_n < \frac{c - \mu_\xi(\lambda_{l_t})}{\sigma_\xi(\lambda_{l_t})}\right\} \geq 1 - \epsilon, \qquad (\text{B.1})$$

where $\mu_\xi(\lambda_{l_t}) = a(\lambda_{l_t})$ and $\sigma_\xi(\lambda_{l_t}) = \left\|\mathbf{b}(\lambda_{l_t})^\top \mathbf{\Sigma}_t^{1/2}\right\|_2$ are the mean and standard deviation of random variable ξ, and the scaled random variable ξ_n has zero mean and unit variance by construction [9].

Assuming ξ as a normal random variable, ξ_n is distributed according to the standard normal distribution. Then, based on the definition of the inverse of the cumulative distribution function of the standard normal distribution Φ, it follows that:

$$\left(\frac{c - \mu_\xi(\lambda_{l_t})}{\sigma_\xi(\lambda_{l_t})}\right) \geq \Phi^{-1}(1 - \epsilon), \qquad (\text{B.2})$$

By substituting the corresponding mean and standard deviation of ξ and rearranging conveniently, (B. 2) results in (3.19).

References

[1] REN21, 2023. Renewables 2023 Global Status Report Collections. [Online]. Available: https://www.ren21.net/gsr-2023

[2] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, *Integrating Renewables in Electricity Markets: Operational Problems*, Springer, 2014.

[3] J. S. Vardakas, N. Zorba, and C. V. Verikoukis, "A Survey on Demand Response Programs in Smart Grids: Pricing Methods and Optimization Algorithms," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 1, pp. 152-178, Jan.-Mar. 2015.

[4] M. Tanaka, A. J. Conejo, and A. S. Siddiqui, *Economics of Power Systems*, Springer, 2022.

[5] A. C. Zambroni de Souza and B. Venkatesh, *Planning and Operation of Active Distribution Networks: Technical, Social and Environmental Aspects*, Springer, 2022.

[6] X. A. Sun and A. J. Conejo, *Robust Optimization in Electric Energy Systems*, Springer, 2021.

[7] A. J. Conejo, M. Carrión, and J. M. Morales, *Decision Making Under Uncertainty in Electricity Markets*, Springer, 2010.

[8] D. Bienstock, M. Chertkov, and S. Harnett, "Chance-Constrained Optimal Power Flow: Risk-Aware Network Control under Uncertainty," *Society for Industrial and Applied Mathematics Review*, vol. 56, no. 3, pp. 461-495, 2014.

[9] L. Roald, F. Oldewurtel, B. Van Parys, and G. Andersson, "Security Constrained Optimal Power Flow with Distributionally Robust Chance Constraints," arXiv:1508.06061v1, Aug. 2015.

[10] "Mensaje de S.E. el Presidente de la República con el que indica un proyecto de ley que establece el derecho a la portabilidad eléctrica," Santiago, Chile. Sep. 4, 2020. [Online]. Available: https://www.camara.cl/

[11] Ley 21.194: Rebaja la rentabilidad de las empresas de distribución y perfecciona el proceso tarifario de distribución eléctrica. Dec. 21, 2019. [Online]. Available: https://www.bcn.cl/leychile/navegar?idNorma=1140301

[12]   Estrategia de flexibilidad para el sistema eléctrico nacional: El camino hacia un sistema eléctrico sostenible. [Online]. Available: https://www.energia.gob.cl/sites/default/files/estrategia_de_flexibilidad.pdf

[13]   Ley 21.505: Promueve el almacenamiento de energía eléctrica y la electromovilidad. Nov. 21, 2022. [Online]. Available: https://www.bcn.cl/leychile/navegar?idNorma=1184572

[14]   X. Chen, Y. Li, J. Shimada, and N. Li, "Online Learning and Distributed Control for Residential Demand Response," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 4843-4853, Nov. 2021.

[15]   J. A. Silva, E. R. Faria, R. C. Barros, E. R. Hruschka, A. C. de Carvalho, and J. Gama, "Data Stream Clustering: A Survey," *ACM Computing Surveys*, vol. 46, no. 1, pp. 1-31, Oct. 2013.

[16]   M. Carnein and H. Trautmann, "Optimizing Data Stream Representation: An Extensive Survey on Stream Clustering Algorithms," *Business & Information Systems Engineering*, vol. 61, pp. 277-297, Jun. 2019.

[17]   J. C. Bezdek and J. M. Keller, "Streaming Data Analysis: Clustering or Classification?," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 91-102, Jan. 2021.

[18]   Z. Jiang, R. Lin, and F. Yang, "An Incremental Clustering Algorithm with Pattern Drift Detection for IoT-Enabled Smart Grid System," *Sensors*, vol. 21, no. 19, pp. 6466, 2021.

[19]   P. Laurinec and M. Lucká, "Interpretable multiple data streams clustering with clipped streams representation for the improvement of electricity consumption forecasting," *Data Mining and Knowledge Discovery*, vol. 33, pp. 413-445, Mar. 2019.

[20]   M. Sun, I. Konstantelos, and G. Strbac, "C-Vine Copula Mixture Model for Clustering of Residential Electrical Load Pattern Data," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2382-2393, May 2017.

[21]   M. Charwand, M. Gitizadeh, P. Siano, G. Chicco, and Z. Moshavash, "Clustering of electrical load patterns and time periods using uncertainty-based multi-level amplitude thresholding," *International Journal of Electrical Power and Energy Systems*, vol. 117, pp. 105624, 2020.

[22]   H. Liang and J. Ma, "Develop Load Shape Dictionary Through Efficient Clustering Based on Elastic Dissimilarity Measure," *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 442-452, Jan. 2021.

[23]   Y. Wang, Q. Chen, C. Kang, and Q. Xia, "Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2437-2447, Sep. 2016.

[24] S. Haben, C. Singleton, and P. Grindrod, "Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 136-144, Jan. 2016.

[25] J. Kwac, J. Flora, and R. Rajagopal, "Lifestyle Segmentation Based on Energy Consumption Data," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2409-2418, Jul. 2018.

[26] F. Lu, X. Cui, J. Xing, S. Liu, Z. Lin, X. Fei, L. Ma, X. Huang, Y. Ding, and L. Yang, "Electricity Load Profile Characterisation for Industrial Users Based on Normal Cloud Model and iCFSFDP Algorithm," *IEEE Transactions on Power Systems*, vol. 38, no. 4, pp. 3799-3813, Jul. 2023.

[27] D. De Silva, X. Yu, D. Alahakoon, and G. Holmes, "A Data Mining Framework for Electricity Consumption Analysis From Meter Data," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 399-407, Aug. 2011.

[28] M. A. Masud, J. Z. Huang, M. Zhong, and X. Fu, "Cluster Survival Model of Concept Drift in Load Profile Data," *IEEE Access*, vol. 6, pp. 51269-51285, Sep. 2018.

[29] G. Le Ray and P. Pinson, "Online adaptive clustering algorithm for load profiling," *Sustainable Energy, Grids and Networks*, vol. 17, pp. 100181, Mar. 2019.

[30] L. Zhao, Z. Chen, Y. Yang, L. Zou, and Z. J. Wang, "ICFS Clustering with Multiple Representatives for Large Data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 3, pp. 728-738, Mar. 2019.

[31] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492-1496, Jun. 2014.

[32] L. E. B. da Silva, N. M. Melton, and D. C. Wunsch, "Incremental Cluster Validity Indices for Online Learning of Hard Partitions: Extensions and Comparative Study," *IEEE Access*, vol. 8, pp. 22025-22047, Jan. 2020.

[33] S. -J. Kim and G. B. Giannakis, "An Online Convex Optimization Approach to Real-Time Energy Pricing for Demand Response," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2784-2793, Nov. 2017.

[34] N. Tucker, A. Moradipari, and M. Alizadeh, "Constrained Thompson Sampling for Real-Time Electricity Pricing With Grid Reliability Constraints," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 4971-4983, Nov. 2020.

[35] Y. Tao, J. Qiu, S. Lai, X. Sun, Y. Ma, and J. Zhao, "Customer-Centered Pricing Strategy Based on Privacy-Preserving Load Disaggregation," *IEEE Transactions on Smart Grid*, vol. 14, no. 5, pp. 3401-3412, Sep. 2023.

[36] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, 2000.

[37]   D. W. Scott, *Multivariate Density Estimation: Theory, Practice and Visualization*, 2nd ed. John Wiley & Sons, 2015.

[38]   S. Guha, R. Rastogi, and K. Shim, "Cure: an efficient clustering algorithm for large databases," *Information Systems*, vol. 26, no. 1, pp. 35-58, Mar. 2001.

[39]   M. Moshtaghi, J. C. Bezdek, S. M. Erfani, C. Leckie, and J. Bailey, "Online cluster validity indices for performance monitoring of streaming data clustering," *International Journal of Intelligent Systems*, vol. 34, no. 4, pp. 541-563, Apr. 2019.

[40]   O. A. Ibrahim, J. M. Keller, and J. C. Bezdek, "Evaluating Evolving Structure in Streaming Data With Modified Dunn's Indices," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 2, pp. 262-273, Apr. 2021.

[41]   Demanda Real. [Online]. Available: https://www.coordinador.cl/operacion/graficos/operacion-real/demanda-real

[42]   D. Tung, H. Trung, and L Bao, "Dynamic Pricing Design for Demand Response Integration in Power Distribution Networks," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3457- 3472, Sep. 2016.

[43]   H. Karimi, R. Bahmani, and S. Jadid, "Stochastic multi-objective optimization to design optimal transactive pricing for dynamic demand response programs: A bi-level fuzzy approach," *International Journal of Electric Power and Energy Systems*, vol. 125, pp. 106487, 2021.

[44]   V. C. Pandey, N. Gupta, K. R. Niazi, A. Swarnkar, and R. A. Thokar, "A Hierarchical Price-Based Demand Response Framework in Distribution Network," *IEEE Transactions on Smart Grid*, vol. 13, no. 2, pp. 1151-1164, Mar. 2022.

[45]   H. J. Monfared and A. Ghasemi, "Retail electricity pricing based on the value of electricity for consumers," *Sustainable Energy, Grid and Networks*, vol. 18, pp. 100205, 2019.

[46]   R. Lu, S. H. Hong, and X. Zhang, "A Dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach," *Applied Energy*, vol. 220, pp. 220, 2018.

[47]   Q. Cai, Q. Xu, J. Qing, G. Shi, and Q. Liang, "Promoting wind and photovoltaics renewable energy integration through demand response: Dynamic pricing mechanism design and economic analysis for smart residential communities," *Energy*, vol. 261, pp. 125293, 2022.

[48]   M. Askeland, T. Burandt, and S. A. Gabriel, "A stochastic MPEC approach for grid tariff design with demand-side flexibility," *Energy Systems*, vol. 14, pp. 707-729, Ago. 2023.

[49]   S. Su, Z. Li, X. Jin, K. Yamashita, M. Xia, and Q. Chen, "Bi-level energy management and pricing for community energy retailer incorporating smart

buildings based on chance-constrained programming," *International Journal of Electric Power and Energy Systems*, vol. 138, pp. 107894, 2022.

[50]   L. He, Y. Liu, and J. Zhang, "An Occupancy-Informed Customized Price Design for Consumers: A Stackelberg Game Approach," *IEEE Transactions on Smart Grid*, vol. 13, no. 3, pp. 1988-1999, May 2022.

[51]   D. Qiu, Y. Wang, J. Wang, C. Jiang, and G. Strbac, "Personalized retail pricing design for smart metering consumers in electricity market," *Applied Energy*, vol. 348, pp. 121545, 2023.

[52]   F. Meng, Q. Ma, Z. Liu, and X. -J. Zeng, "Multiple dynamic pricing for demand response with adaptive clustering-based customer segmentation in smart grids," *Applied Energy*, vol. 333, pp. 120626, 2023.

[53]   S. A. Gabriel, A. J. Conejo, J. D. Fuller, B. F. Hobbs, and C. Ruiz, *Complementarity Modeling in Energy Markets*, Springer, 2013.

[54]   E. Dall'Anese, K. Baker, and T. Summers, "Chance-Constrained AC Optimal Power Flow for Distribution Systems With Renewables," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3427-3438, Sep. 2017.

[55]   R. Mieth and Y. Dvorkin, "Data-Driven Distributionally Robust Optimal Power Flow for Distribution Systems," *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 363-368, Jul. 2018.

[56]   R. Mieth and Y. Dvorkin, "Distribution Electricity Pricing Under Uncertainty," *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 2325-2338, May 2020.

[57]   R. Mieth and Y. Dvorkin, "Online Learning for Network Constrained Demand Response Pricing in Distribution Systems," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2563-2575, May 2020.

[58]   S. Zhang, S. Ge, H. Liu, J. Li, C. Gu, and C. Wang, "A Multi-objective Chance-constrained Information-gap Decision Model for Active Management to Accommodate Multiple Uncertainties in Distribution Networks," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 1, pp. 17-34, Jan. 2023.

[59]   K. S. Ayyagari, R. Gonzalez, Y. Jin, M. Alamaniotis, S. Ahmed, and N. Gatsis, "Learning Reactive Power Control Policies in Distribution Networks Using Conditional Value-at-Risk and Artificial Neural Networks," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 1, pp. 201-211, Jan. 2023.

[60]   L. Yang, Y. Xu, H. Sun, and W. Wu, "Tractable Convex Approximations for Distributionally Robust Joint Chance-Constrained Optimal Power Flow Under Uncertainty," *IEEE Transactions on Power Systems*, vol. 37, no. 3, pp. 1927-1941, May 2022.

[61]   W. Xie, S. Ahmed, and R. Jiang, "Optimized Bonferroni approximations of distributionally robust joint chance constraints," *Mathematical Programming*, vol. 191, pp. 79-112, Nov. 2019.

[62]   S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Mathematical Programming*, vol. 137, pp. 167-198, Nov. 2011.

[63]   M. Lubin, Y. Dvorkin, and L. Roald, "Chance Constraints for Improving the Security of AC Optimal Power Flow," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 1908-1917, May 2019.

[64]   M. E. Baran and F. F. Wu, "Optimal capacitor placement on radial distribution systems," *IEEE Transactions on Power Delivery*, vol. 4, no. 1, pp. 725-734, Jan. 1989.

[65]   M. S. Nazir, I. A. Hiskens, A. Bernstein, and E. Dall'Anese, "Inner Approximation of Minkowski Sums: A Union-Based Approach and Applications to Aggregated Energy Resources," arXiv:1810.01587v1, Oct. 2018.

[66]   R. Sioshansi and A. J. Conejo, *Optimization in Engineering: Models and Algorithms*, Springer, 2017.

[67]   J. Fortuny-Amat and B. McCarl, "A Representation and Economic Interpretation of a Two-Level Programming Problem," *Journal of the Operational Research Society*, vol. 32, pp. 783-792, Sep. 1981.

[68]   YALMIP. [Online]. Available: https://yalmip.github.io

[69]   IEEE 37 Node Test Feeder. [Online]. Available: https://cmte.ieee.org/pes-testfeeders/resources

[70]   A. Molina, M. Falvey, and R. Rondanelli, "A solar radiation database for Chile," *Scientific Reports*, vol. 7, pp. 14823, Nov. 2017.

[71]   IEEE Standard Power Cable Ampacity Tables, in *IEEE Std 835-1994* (R2012), Dec. 1994.

[72]   Q. Nguyen, H. V. Padullaparti, K. -W. Lao, S. Santoso, X. Ke, and N. Samaan, "Exact Optimal Power Dispatch in Unbalanced Distribution Systems With High PV Penetration," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 718-728, Jan. 2019.

[73]   M. Rayati, M. Bozorg, R. Cherkaoui, and M. Carpita, "Distributionally Robust Chance Constrained Optimization for Providing Flexibility in an Active Distribution Network," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 2920-2934, Jul. 2022.

[74]   T. Ding, C. Li, Y. Yang, J. Jiang, Z. Bie, and F. Blaabjerg, "A Two-Stage Robust Optimization for Centralized-Optimal Dispatch of Photovoltaic Inverters

in Active Distribution Networks," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 2, pp. 744-754, Apr. 2017.

[75]   F. M. Aboshady, I. Pisica, A. F. Zobaa, G. A. Taylor, O. Ceylan, and A. Ozdemir, "Reactive Power Control of PV Inverters in Active Distribution Grids With High PV Penetration," *IEEE Access*, vol. 11, pp. 81477-81496, Aug. 2023.

[76]   IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces, in *IEEE Std 1547-2018*, Feb. 2018.