2017

Online monitoring and control of voltage stability margin via machine learning-based adaptive approaches

Shiyang Li
Iowa State University

Online monitoring and control of voltage stability margin via machine learning-based adaptive approaches

by

Shiyang Li

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Electrical Engineering (Electric Power and Energy Systems)

Program of Study Committee:
Venkataramana Ajjarapu, Major Professor
Zhaoyu Wang
Umesh Vaidya
Chinmay Hegde
Kris De Brabanter

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

DEDICATION

I would like to dedicate this thesis to my grandparents who passed away during my Ph.D. years.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

ACKNOWLEDGEMENTS

Following the tradition, I should take this opportunity to express my thanks to those who helped me with various aspects of conducting research in the past five years. However, I find this is actually inapplicable, because the list of names and thanks will be definitely as long as a major chapter, or, instead, I need to create a new chapter to develop the proper "feature selection" algorithm that extracts the most effective contributors in my Ph.D. learning process. So, I decide to only mention the names of those who are directly related to this dissertation.

First, I would like to express my sincere appreciation and gratitude to my advisor Prof. Venkataramana Ajjarapu for his valuable guidance and support throughout the course of my doctoral studies. He just showed me what a true scholar should be. He taught me how to communicate with others (discussing, presenting, writing a paper or a report, etc.): (i) being active, (ii) trade-off between understandability and accuracy, (iii) converging in iterations, and (iv) using others' words. Suddenly I find these principles are just the spirit of my proposed approaches in this dissertation.

My deepest gratitude is due to my POS committee members for their valuable time and comments. I wish to express my special thanks to Prof. Kris De Brabanter and Prof. Chinmay Hegde. In the initial stage of my research, Dr. De Brabanter suggested to consider using LASSO for the regression task, instead of directly using local regression for high-dimensional problem. This finally inspired me to combined them. In my preliminary exam and final defense, he pointed out some important defects in my original work and provided constructive ideas, which helped me to improve the statistical strictness of the work and get some ideas for future researches. Dr. Hegde carefully reviewed my dissertation draft and provided very useful comments. Many basic concepts of machine learning I got from the courses of these two professors.

My friend Yihua Li carefully proofread the major part the dissertation and checked all mathematic derivations. More importantly, many critical ideas of this work came out of the discussions

between us, including observing the relative regularization factor from Bayesian point of view, shrinking the effective domain of control action via PCA, equally weighting the neighbors in the local regression for control, robustly correcting the scaling factors of VSM predictive model coefficients, explaining my inference about the VSM prediction as prediction interval, and so on. All these ideas resolved the toughest problems in my final work. I would like to thank Yihua for her patience, encouragement, and for always lighting up my wisdom.

I would also like to express my special thanks to Shanshan Ma for her kindly care, support, and dissuasion during the most stressful stage of writing the dissertation. Without her help, a smooth oral defense is impossible.

Zixiao Ma helped me to clarify my thoughts about using robust model predictive control. Although finally that is not realized, it helps me to investigate control stability in my problem.

I would additionally like to thank my fellow students in our voltage stability group: Magesh Paramasivam, Nicholas Brown, A. Ramapuram-Matavalam, Ashraf Radaideh, Ankit Singhal, Pranav Sharma, Qian Zhang, Ramakrishnan Venkatraman, and Alok Bharati. You made me confident in the preliminary exam and the final defense, because nearly every slide and every word had survived your repeated scrutinies. We always work as a team.

I wish to express my special thanks to my families for their amazing support and scarifies. My parents came to accompany at the most difficult time of creating the control test example, and left in the final stage to let me concentrate. I know sometimes you are even more stressed than me, but you are always confident and proud of me. Words cannot express how grateful I am to everything you did.

Finally, I would like to thank this thrilling new era of smart grid and AI, where anxieties and opportunities spring up every day. At least, it would be much more difficult to convince people of the potential of my research, if I started my Ph.D. just five years earlier.

ABSTRACT

Voltage instability or voltage collapse, observed in many blackout events, poses a significant threat to power system reliability. To prevent voltage collapse, the countermeasures suggested by the post analyses of the blackouts usually include the adoption of better online voltage stability monitoring and control tools. Recently, the variability and uncertainty imposed by the increasing penetration of renewable energy further magnifies this need. This work investigates the methodologies for online voltage stability margin (VSM) monitoring and control in the new era of smart grid and big data. It unleashes the value of online measurements and leverages the fruitful results in machine learning and demand response.

An online VSM monitoring approach based on local regression and adaptive database is proposed. Considering the increasing variability and uncertainty of power system operation, this approach utilizes the locality of underlying pattern between VSM and reactive power reserve (RPR), and can adapt to the changing condition of system. LASSO (Least Absolute Shrinkage and Selection Operator) is tailored to solve the local regression problem so as to mitigate the curse of dimensionality for large-scale system. Along with the VSM prediction, its prediction interval is also estimated simultaneously in a simple but effective way, and utilized as an evidence to trigger the database updating. IEEE 30-bus system and a 60,000-bus large system are used to test and demonstrate the proposed approach. The results show that the proposed approach can be successfully employed in online voltage stability monitoring for real size systems, and the adaptivity of model and data endows the proposed approach with the advantage in the circumstances where large and unforeseen changes of system condition are inevitable.

In case degenerative system conditions are identified, a control strategy is needed to steer the system back to security. A model predictive control (MPC) based framework is proposed to maintain VSM in near-real-time while minimizing the control cost. VSM is locally modeled as a

linear function of RPRs based on the VSM monitoring tool, which convexifies the intricate VSM-constrained optimization problem. Thermostatically controlled loads (TCLs) are utilized through a demand response (DR) aggregator as the efficient measure to enhance voltage stability. For such an advanced application of the energy management system (EMS), plug-and-play is a necessary feature that makes the new controller really applicable in a cooperative operating environment. In this work, the cooperation is realized by a predictive interface strategy, which predicts the behaviors of relevant controllers using the simple models declared and updated by those controllers. In particular, the customer dissatisfaction, defined as the cumulative discomfort caused by DR, is explicitly constrained in respect of customers' interests. This constraint maintains the applicability of the control. IEEE 30-bus system is used to demonstrate the proposed control strategy.

Adaptivity and proactivity lie at the heart of the proposed approach. By making full use of real-time information, the proposed approach is competent at the task of VSM monitoring and control in a non-stationary and uncertain operating environment.

CHAPTER 1. OVERVIEW

1.1 Introduction

1.1.1 Voltage Stability

Modern power system is more likely to be heavily stressed due to the trends of (i) large-area network interconnection and power transfer, (ii) market-oriented deregulation, and (iii) integration of renewable energy that is sometimes far away from load centers. As a result, power system voltage stability, which inherently connects to the loadability of the system, became the research topic of interest for the past several decades [1, 2, 3]. IEEE/CIGRE joint task force proposed various definitions related to power system stability including voltage stability [4]. Figure 1.1 summarizes the classification of power system stability. In general terms, *voltage stability* is defined as "*the ability of a power system to maintain steady voltages at all buses in the system after being subjected to a disturbance from a given initial operating condition. It depends on the ability to maintain/restore equilibrium between load demand and load supply from the power system. Instability that may result occurs in the form of a progressive fall or rise of voltages of some buses*" [4].

A possible outcome of voltage instability is loss of load in an area, i.e., a blackout. The term *voltage collapse* is used to describe "the process by which the sequence of events accompanying voltage instability leads to a blackout or abnormally low voltages in a significant part of the power system" [4]. Many large blackout events in history are mainly or partially attributed to voltage instability or voltage collapse [5, appendix. F], [6, chap. 1], including the Northeast blackout in 2003 that impacted more than 50 million people [7]. Thus, to prevent the catastrophic blackouts caused by voltage collapse, theories and techniques for voltage stability assessment and control are consistently among the critical research tasks of power community since 1980s.

This work focuses on the study of *long-term voltage stability*. It considers the stability in the long run, typically as loads slowly increase under various operating conditions. Thus, it could

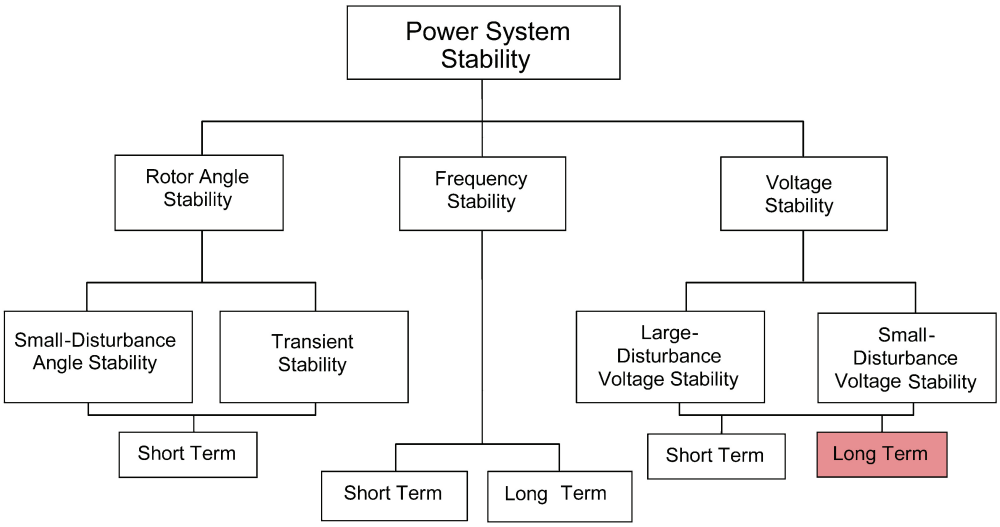


Figure 1.1: Classification of power system stability [4].

involve slower acting equipment such as tap-changing transformers, thermostatically controlled loads, and generator current limiters. Based on the connection between saddle-node bifurcation (SNB)/limit-reduced bifurcation (LIB) and the steady-state loading limit [8, 9, 10, 11], this work specifically considers the long-term dynamic problem that can be properly simplified and analyzed using algebraic equilibrium equations, which is mainly recognized as a small-disturbance voltage stability problem.

1.1.2 Model Simplification for Long-Term Voltage Stability Analysis

A research focus implies the specific modeling. Power system is a nonlinear dynamic system that generally can be formulated as ordinary differential equations (ODE). Based on the theory of singular perturbation and time-scale decomposition of the dynamics [12], the dynamics faster than what we concern (e.g. electromagnetic dynamics) can be ignored under proper conditions by their equilibrium equations, and the dynamics slower than what we concern (e.g. sustained load buildup) can be ignored by fixing the state variables as parameters. Thus, for certain study purpose, power system generally can be simplified as a set of differential algebraic equations (DAE) [13]. (1.1) shows this simplification process, where s denotes the multi-dimensional state variable

of the power system, and it can be decomposed into three components x, y, and z which denote the state variables for the concerned dynamics, fast dynamics, and slow dynamics respectively; F gives the full representation of the nonlinear dynamics of the system, which can be transferred to three components f, g, and h corresponding to x, y, and z; z is regarded as a given parameter after simplification.

$$\textcircled{1}:\ \dot{s} = F(s) \xrightarrow[\text{decomposition}]{\text{time-scale}} \textcircled{2}: \begin{cases} \dot{z} = h(x,y,z) \\ \dot{x} = f(x,y,z) \\ \dot{y} = g(x,y,z) \end{cases} \xrightarrow{\text{simplificatoin}} \textcircled{3}: \begin{cases} \dot{x} = f(x,y,z) \\ 0 = g(x,y,z) \end{cases} \tag{1.1}$$

Under the causality condition [9], $0 = g(x,y,z)$ determines an implicit function from (x,z) to y, then substitute it into f gives the explicit ODE of the simplified system ③:

$$\dot{x} = f_d(x,z) = f(x, y(x,z), z) \tag{1.2}$$

The system is said to undergo a *bifurcation* at $z = z_c$ if the flow (informally speaking, the solution phase portrait of x) of (1.2) at $z = z_c$ is not topologically equivalent to the flow for z near z_c [14]. In particular, a *saddle-node bifurcation* (SNB) occurs when two equilibrium points of (1.2) collide and annihilate each other as z changes across z_c along certain path. In the neighborhood of z_c along that path, a stable invariant manifold in the state space of x becomes unstable, thus, SNB implies the instability in the Lyapunov sense. This has been recognized as a major mechanism of voltage collapse [9].

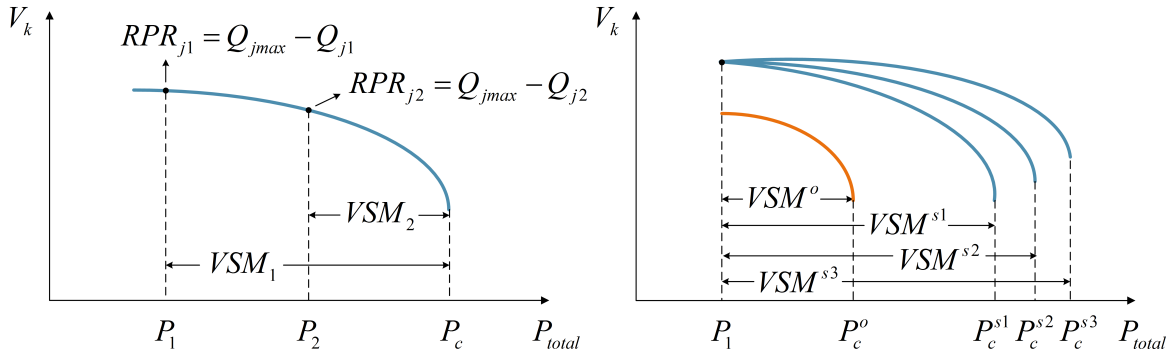
A critical necessary condition of SNB is that the Jacobian of f_d is singular. This can be indicated by that the Jacobian of (1.1)③ (i.e. $\frac{\partial(f,g)}{\partial(x,y)}$) is singular, then further indicated by the singularity of the standard power flow equations when the bus type assumptions (PV, PQ, Vθ) held well [8, 9, 10, 11]. These singularities can also be obtained via the K.K.T. conditions for maximizing the change of z along certain direction with the power flow equations, or the equilibrium equations of (1.1)③ or (1.2), as constraints. Thus, in case z is the load, SNB, as the long-term voltage stability limit point, is approximated by the steady-state loading limit, and arguably can be analyzed using the algebraic equilibrium equations, or simply power flow equations.

1.1.3 Indices for Long-Term Voltage Stability

Given the simplified model mentioned above, usually just power flow equations, a major task of voltage stability analysis is voltage stability assessment. Numerous indices have been constructed to measure the long-term voltage stability. Determinant, smallest eigenvalue, smallest singular value of the Jacobian, and certain sensitivity of voltage with respect to power injection, are inherent indices by directly indicating the singularity of the Jacobian. However, they do not have a normalized range, thus cannot be used for comparing stability among different systems. Other indices which also can be calculated using current state (or a few nearby states in the case of impendence matching index based on Thévenin equivalent) are constructed to have a normalized range, typically between zero and one. Definitions and comparisons of these indices can be found in [15, 16, 17, 3, 18, 19].

The sign of these indices effectively indicates the singularity of Jacobian as a sign of SNB, and their magnitude can monotonically reflect the extent of stability in some sense. However, the exact value of these indices usually does not has clear physical meaning, so are rarely used to guide the system operation. On the contrary, we can define a index using certain metric in the parameter space, which measures the physically meaningful distance between the current operating point and the SNB. This type of indices is recognized as *voltage stability margin* (VSM). For example, when the parameter is load power, the *load margin* tells how much load increment is still affordable without voltage collapse, based on the current operating point and the *hypothetical scenario* of load increase. In fact, this is a natural choice for voltage stability index, given that we have established the connection between collapse point and steady-state loading limit (see 1.1.2). From the perspective of online voltage stability monitoring, margin index is a more intuitive description about how far the system is from voltage collapse, thus enhances the situation awareness of operators. Moreover, since the parameters used to define a margin (e.g. the load) are usually controllable, margin index also provides actionable information that controls can rely on. With these advantages, load margin is one of the most widely used voltage stability index that has been applied in many aspects of power system planing and operation. In this paper, by means of PV curve, we defined the VSMas

the distance between the current operating point and the loading limit point in a hypothetical scenario, measured by the total active power increment, see Figure 1.2.



(a) Different base operating points. (b) Different hypothetical scenarios.

Figure 1.2: Definition of voltage stability margin (VSM) and reactive power reserve (RPR). (a) Two instances for different base points under the same hypothetical scenario are annotated. (b) Four PV curves under different hypothetical scenarios are drawn. The blue ones correspond to three distinct load increase directions but the same base point; the orange one corresponds to the PV curve after a contingency such as an outage of generator or transmission line.

1.1.4 Reactive Power Reserve

Reactive power reserve (RPR) of a reactive power source refers to its available reactive power generating capability. In other words, the RPR of a reactive power source j is defined by the difference between its maximal reactive generation Q_j^{max} and the current reactive generation Q_j (see Figure 1.2a). [20] investigated several possible setting of Q_j^{max}. In this work we simply use constant Q_j^{max}, which is the standard setting for power flow problem.

A basic factor for power transmission network is that on normal operating conditions, bus voltage has a strong connection with reactive power injection, but relatively weak connection with active power injection. Therefore, RPR is naturally regarded as the available capability of the system to maintain voltage level, which inherently connects to voltage stability, and insufficient RPR is a common reason in many blackout events [7]. Thus, reactive power source planing and RPR management is among the major tasks of today's utility to maintain the power system reliability.

The North American Electric Reliability Corporation (NERC) has mandates several standards requiring real-time monitoring of RPR [21, 22].

1.1.5 Demand Response

The Federal Energy Regulatory Commission (FERC) in order 755 has motivated the utilities to search for additional resources to support the system flexibility and invest more in clean resources [23]. Demand Response (DR) resources have been introduced recently for utilizing load-flexibility at the aggregate level and benefit the overall system needs. DOE (Department of Energy of the U.S.) provides a brief introduction of DR in its website [24]:

"Demand response provides an opportunity for consumers to play a significant role in the operation of the electric grid by reducing or shifting their electricity usage during peak periods in response to time-based rates or other forms of financial incentives. Demand response programs are being used by some electric system planners and operators as resource options for balancing supply and demand. Such programs can lower the cost of electricity in wholesale markets, and in turn, lead to lower retail rates. Methods of engaging customers in demand response efforts include offering time-based rates such as time-of-use pricing, critical peak pricing, variable peak pricing, real time pricing, and critical peak rebates. It also includes direct load control programs which provide the ability for power companies to cycle air conditioners and water heaters on and off during periods of peak demand in exchange for a financial incentive and lower electric bills. The electric power industry considers demand response programs as an increasingly valuable resource option whose capabilities and potential impacts are expanded by grid modernization efforts."

Electric loads once operating and consuming power are considered synchronized to the power systems and their contribution can provide potential support to the system in various time-scales. Reserves extracted from flexible loads are equivalent to contributions from other thermal units, and the aggregated response from DR programs can be more valuable and economically feasible than services provided by other peaking units. Therefore, DR resources are counted as a potential source for future grid flexibility [25].

1.1.6 The Need, Challenge, and Opportunity for Real-Time Margin Monitoring and Control

The NERC standard [26] requires to establish the System Operating Limits (SOLs) methodology for the operations horizon, by which "*the system shall demonstrate transient, dynamic and voltage stability; all Facilities shall be operating within their Facility Ratings and within their thermal, voltage and stability limits*", under normal condition or N-1 contingencies. Although monitoring the online state measurements (e.g. RPRs, voltages, power flows, etc.) involved in SOLs may enhance operators awareness regarding whether the operating state is secure or not, and it is possible to enhance stability through simply steering the system away from SOLs (e.g. RPR management may improve voltage stability [27, 28]), the raw measurements cannot provide quantitative information like VSM to show how far the system is from a voltage collapse. The Federal Energy Regulatory Commission (FERC) has questioned the efficacy of the some NERC standards by saying that system operators cannot gain situational awareness by simply viewing massive amounts of raw data. According to [29], " *while the requirements identify the data to be gathered, they fail to describe the tools necessary to turn that data into critical reliability parameters*". The post analyses of many blackout events have shown the necessity and benefits of adoption better stability monitoring and control tools [7]. Therefore, there is a need to develop tools that can transfer the massive amount of data gathered in the energy management system (EMS) into meaningful information that indicates voltage stability (e.g. VSM) in real time. Then, an online control strategy is needed to steer the system back to security once low VSM is observed.

A major challenge for this objective is that VSM is computationally intractable in the real-time horizon. Classical methods for exact VSM evaluation, i.e. continuation power flow [30, 31, 32], direct method [33, 34, 35] and nonlinear optimization [36, 37] are quite computationally intensive for detailed model of large system, especially when multiple hypothetical scenarios are considered. Thus, VSM evaluation based on these methods usually cannot be finished in real time, let alone used in a VSM constrained optimal control where multiple evaluations of VSM is inevitable during the iteration.

The big data techniques, especially the fruitful achievements in *machine learning*, provide great opportunities to quickly make sense of the online measurements. By this class of methods, first, sufficiently many VSMs of historical and simulated operating points are calculated using aforementioned classical methods. Then a statistical predictive model is trained from the state measurements and corresponding VSMs of these operating points using machine learning techniques. During the training process, the predictive model learns the underlying relationship between the measurements and VSM, implied by the system model mentioned in 1.1. That yields an explicit representation of VSM as a function of the measurements. In this way, VSM can be evaluated in-real time for online operating condition, and used to construct VSM constraint in an optimization problem.

1.1.7 The Need for Adaptive Approach

Efforts have been made in developing machine learning-based method for VSM online monitoring and control (see Chapter 2). However, most of the existing monitoring methods, mainly based on off-line training, tend to underestimate the complexity and variability of the underlying pattern between the measurements and VSM (detailed in 3.1.6), and underestimate the variability and uncertainty of the control environment (i.e. state and availability of control resources, impacts of other controllers, etc.). Consequently, they cannot well adapt to the changing condition of power system operation. Thus, there is a need to investigate the adaptive methodology for online VSM monitoring and control.

Recently, this need is further magnified due to the variability and uncertainty imposed by the increasing penetration of renewable energy. On-off switching of units including generators, which structurally changes the system equations mentioned in 1.1, becomes much more frequent. In such circumstances, a permanent or periodically updated off-line predictive model cannot follows the operating condition, thus may lead to unreliable predictions and controls. This trend of power system development further motivate us to develop an adaptive approach for online VSM monitoring and control in such a non-stationary operating environment.

1.2 Research Objective

The objective of this work is to develop a comprehensive methodology for online long-term voltage stability monitoring and control in the new era of smart grid and big data. This approach should be valid in an uncertain and non-stationary operating environment. In contrast to previous approaches, this work resolves and adapts to the uncertainty by fully leveraging the online information in addition to the off-line knowledge, and cooperatively engages in the demand response as a new resource for stability control. The focus is on developing study methodology, procedures and tools to support transmission operators in real-time voltage stability assessment, and in preventive stability control under emergency.

1.3 Organization of Dissertation

This dissertation includes two major parts: (i) real-time voltage stability margin (VSM) monitoring via local regression and adaptive database, (ii) MPC-based online VSM enhancement. The rest of the dissertation is organized as follows:

Chapter 2 reviewed the important literature in the scope of online voltage stability assessment and online VSM enhancement methodology, focusing on the relevant works using machine learning and MPC.

Chapter 3 proposed the real-time VSM monitoring tool using local regression and adaptive database. The motivation of developing an adaptive approach was thoroughly discussed. Then, the two technical blocks - local regression and adaptive database are presented in detail. Finally, it is summarized and demonstrated on IEEE 30-bus system and a real large-scale system.

Chapter 4 proposed the MPC-based online VSM enhancement approach. The requirements for the concerned control approach were discussed at first. Then, the key techniques for modeling VSM and integrating demand response were described in detail. The overall formulation and application procedure were summarized, followed by a test example on IEEE 30-bus system.

Chapter 5 presents the final conclusions and contributions of this work, and discusses possible future works.

CHAPTER 2. REVIEW OF LITERATURE

2.1 Online Assessment of Voltage Stability

[38] provided a survey on existing online voltage stability assessment (VSA) methods up to 2011. It classified VSA methods into two broad categories according to the inputs: (i) methods based on spacially local measurements, which are usually used to support fast local preventive or corrective controls; and (ii) methods requiring the observability of the whole region (or a large area) prone to voltage instability, which offer the potential advantages of wide-area monitoring, and they are usually used to support preventive controls on the system level. In this section, VSA methods are clustered into five categories according to the methodology and the specific target, followed by an in-depth review for the machine learning-based methods.

2.1.1 Major Approaches

2.1.1.1 Methods using model-based state space indices

As mentioned in 1.1.3, a bunch of indices have been constructed to indicate voltage instability or voltage collapse [15, 16, 17, 3, 18, 19]. Some of them are defined in state space (as a function of online state measurements such as voltages, currents and power injections), whereas others are defined in parameter space (as a metric on parameter space such as load space). The latter are usually called voltage stability margin (VSM) which provides physically meaningful information about the distance to voltage collapse. Although, according to the analyses on power flow equations, VSM could also be approximated using state variables [39, 40, 41, 42], sometimes indirectly through state space indices, here it is regarded as a separate type of indices whose exact values do not have explicit expressions.

Most of the state space indices are based on power flow model or its simplification, while some others, such as Thévenin-based indices, can be evaluated in a model-free procedure. Here we regard

them as two different categories, and Thévenin-based methods will be described in 2.1.1.2. Wide-area measurements (centralized or distributed [43, 44, 45]) or local measurements could be used as the inputs for this category. Besides SCADA, phasor measurement units (PMUs) are also utilized to provide the inputs for index evaluation [46, 47, 48, 49, 44, 45].

This category of methods use online measurements of one operating state as inputs (special cases are the methods based on the multi-solution property of power flow [50, 51, 52], which also need the conjugate low voltage power flow solution). Typically, they are used for quickly indicating the onset of voltage instability, triggering a warning for marginally stable operating condition, or approximating VSM based on full/simplified/local power flow model, via evaluation of explicit function of the inputs. However, most of them cannot provide margin information. Some of them can approximate the margin but need strong assumptions or simplifications of the system.

2.1.1.2 Methods based on Thévenin equivalent

This category of methods [15, 53, 54, 55, 56, 48, 57, 58, 59, 60, 61, 62] are based on the simple principle from circuit theory: the output load power of an independent linear AC network reaches its maximum when the load impedance matches the Thévenin equivalent impedance of the network (two impedances have equal magnitude). Thus, voltage stability can be indicated by the ratio of the two impedances. The key of these methods is that the two impedances, or their ratio directly, can be simply estimated by online measurements without any model parameters, thus this category of methods could be model-free. For supporting fast local control, they can use only local measurements to get the single port impedance ratio as a local instability indicator. Or the ratios from multiple ports can be assembled at the control center as a wide-area indicator.

A model-based variation for this category introduces the concept of coupled single-port circuit model [58, 59, 60, 61, 62]. Based on the linear network equations (of voltages and currents), the port character can be equivalently described by an extended Thévenin circuit which includes an extra component (source, load, or impedance) to reflect the coupling with unseen independent variables (e.g. current injections at other buses). Comparing to classical Thévenin equivalent, the coupled

version utilized wide-area information more deeply and could be used to approximate VSM under proper conditions. However, it requires wide-area measurements and network model.

[59] compared some of aforementioned methods by time domain simulation.

There are also semi-model-based variations for this category, where system model is used to reduce the need of measurements at two distinct operating points to that of only one [63], or used to validate and calibrate the impedance estimates from the measurements [64]. Another benefit of combining model and measurements is that it is usually less sensitive to the measurement noise.

Research on this category is pretty active due to the model-free property and the implementation of PMU. However, there are some inherent challenges that may limit these methods: (i) it is sensitive to the quality of the measurements, so some filters are needed to smooth the measurements while capture the actual moving of the steady state; (ii) the two instants of measurements cannot be too close (otherwise noise will be a big issue, and the equations for impedance estimation will be ill-conditioned), while they also cannot be too far away (otherwise they correspond to different operating points which do not have the same Thévenin parameters), so a subtle strategy is needed to choose the sampling interval; (iii) the approximation of VSM from Thévenin equivalent, if any, is usually under strong assumptions and simplifications.

2.1.1.3 Methods based on PV curve tracing

PV curve tracing, via continuation power flow [30, 31, 32] or consecutively solving power flow, is the most reliable static way to calculate VSM since it can consider the detailed model of system which usually has very complicated (discrete, of composite logic, path-dependent) behaviors during load increase. This advantage itself implies its limitation: it is quite computational intensive and deeply model-dependent. Thus, for large system, it cannot provide VSM evaluation in real time. For online application, usually it is used in the study mode of EMS, in order to validate VSM prediction from other fast methods, or provide detailed analyses on selected critical contingencies. As a classic method to calculate VSM, it has been widely commercialized in many ESM applications [65, 66].

Numerous efforts have been made to speed-up the process of PV curve tracing, among which, Dimo's method is reported to achieve real-time level for practical system. This method was developed by Paul Dimo [67, 68, 69, 70], and successfully implemented in several running SCADA/EMS installations [69]. It follows the same idea of stepwise stressing the system until the voltage collapse, but the stressing process is realized by a "case-worsening procedure" instead of performing a sequence of power flow computations. The fast case-worsening procedure is enabled by (i) a strong simplification of system including short-circuit current network transformation and aggregated fictitious load center, and (ii) using $\Delta Q/\Delta V$ as collapse point indication, which leads to a reduction in accuracy.

2.1.1.4 Methods based on voltage stability region

The idea of this category is, if the entire boundary of voltage stable operating region in parameter space [71, 72, 73, 74] can be explicitly represented off-line, then online VSM evaluation is simply calculating the geometrical distance between the current operating point and the boundary. Many techniques have been developed to obtain the boundary through sampling-fitting or machine learning [75, 76, 77, 78] (note: although machine learning techniques like neural networks are used, these works target at the analytical representation of the security or stability boundary rather than predictive model of VSM, so they are regarded as region based methods rather than machine learning based methods described in the next section), tangent hypersurface [79, 80, 81], or function approximation for bifurcation equations [82]. When the concerned parameter space has only 2 or 3 dimensions, continuation methods also be used to trace the boundary [6].

Many utilities' control area can be divided into a small number of generation or load centers, and the operators mostly concern the transfer limits among these centers as part of the system operating limits (SOLs). To this end, the methods of this category can be used to visualize the voltage stability boundary in the transfer power space, so as to enhance the situation awareness. However, the scalability is a big issue for these methods, because all of them inherently face the curse of dimensionality in representing a hypersurface in high dimensional parameter space.

2.1.1.5 Methods based on machine learning

From 2.1.1.1 to 2.1.1.4 one can clearly see that the trade-off between evaluation efficiency and model accuracy is a key problem in designing a VSM monitoring method. To achieve real-time level efficiency, all the aforementioned methods, except the one approximating stability boundary by sampling-fitting, simplify the underlying model by *deduction* using power system domain knowledge. Another possibility is approximating the underlying model by *induction* from data. Methods based on machine learning can be helpful in this area.

First, these methods evaluate certain voltage stability measure (e.g. get VSM by tracing PV curves) under various operating conditions and hypothetical scenarios using the detailed system model (even using real-time simulation of dynamic model). Thus, a database of online state measurements and corresponding voltage stability measures is established. The state measurements are the ones that can be directly obtained from SCADA/PMU/EMS system, such as voltages, reserves of generators, and power flows. The voltage stability measure could be a continuous index mentioned in 1.1.3, or simply a categorical description such as "stable", "unstable", "emergent". Then, a statistical predictive model that directly maps the state measurements into voltage stability measure is trained from the database using *supervised machine learning* techniques. During the training process, the predictive model learns the underlying relationship between the state measurements and the voltage stability measure, and finally reaches an explicit representation of the voltage stability measure as a function of the state measurements, which can be quickly evaluated for online operating point in real time.

In contrast to the underlying models adopted in most methods of other categories, the predictive model here is obtained through a synthetic procedure. Based on statistics, these methods inherently consider the uncertainties of operating point and hypothetical scenarios (typically, it's output is a mathematical expectation), which makes them preferable in an uncertain operating environment. Besides, compared to the methods of other categories, machine-learning based methods are concurrently endowed with the favorable features that they are (i) based on detailed system

model, (ii) applicable in real time, (iii) scalable for large system, and (iv) applicable for what-if scenario. The techniques in this category will be reviewed in detail in the next section.

2.1.2 Machine Learning-Based Voltage Stability Monitoring

Various learning techniques have been investigated for online voltage assessment. The outputs of these predictive models could be real-valued VSM (regression), or categorical security level (classification) [83]. Most of the literature in this field are based on artificial neural network (ANN) [84, 85, 86, 83, 87, 88]. The general comment is that ANN usually achieves superior in-sample accuracy, and proper tuning techniques could also release its potential to gain a better performance in predicting new data [89, 90].

However, due to its complex and nonlinear model structure, the training process of ANN is usually computationally cumbersome and the mapping between input and output is opaque (black/grey box without a simple closed form). Notice that such opaqueness of ANN sometimes is not acceptable when it is used in safety critical applications like power system stability monitoring and control. Efforts are devoted to extracting closed-form rules from ANN to reveal the black box [91, 83]. However, this could be done only after the ANNs are obtained, which further aggravates the computational burden. Besides, in the literature mentioned above, each ANN is usually trained for particular network topology, indicating that the number of ANNs could grow out of proportion in case hundreds of topologies are considered, and the final prediction accuracy relies on online topology identification. [92] showed the possibility of using only one neural network for different system topologies, but only a very small number of most severe outages are considered in the applications.

Except for ANN, other learning techniques such as decision tree (DT) [93, 94, 95] , random forest [96], support vector machine (SVM) [97], bagging [98] and linear regression [99, 100] are also applied for VSA. [98] compared these methods and suggested that an additive and transparent regression model is more robust to missing inputs and can well balance the accuracy and simplicity/transparency, which could be crucial for security sensitive problem like stability monitoring.

Along with learning algorithm, the problem of input feature selection is another focus of research. The inputs of the statistical models mentioned above are usually first selected by heuristics. Load looks like intuitively the first choice because the load margin itself is measured by the load change. Besides, generations and reserves of generators, power flows and losses in transmission lines, and voltages are usually among the candidates. The set of such input features can be further reduced either based on experience or the mathematical correlations between a feature and VSM. For example, the critical load bus or generator location methods developed for voltage stability analysis and control can be utilized to for feature selection. Besides, standard techniques for feature selection and dimension reduction, such as principle component analysis (PCA), provide abundant tools to identify or extract effective features.

In this work, we emphasize the special role of RPR as input feature. As mentioned in 1.1.4, RPR is inherently connected to VSM, and therefore NERC has issued standards related to RPR real-time monitoring [21, 22]. Also, RPR has some favorable features which will be discussed in 3.1.3.

Taking advantage of the simplicity and transparency of linear models, [99] explored the RPRs as potential indicators of VSM. In this work, an online voltage stability monitoring system using the concept of equivalent RPR is proposed, where VSM is predicted by a linear function of equivalent RPRs. Results have shown that RPRs can be effective indicators of VSM.

[100] extended this idea. Multi-linear regression model (MLRM) with quadratic terms were used to model the relationship between VSM and RPRs. Operating conditions are sampled off-line over a large range of contingencies, operating scenarios and load increase directions, and then divided into a few groups. MLRMs are trained for each group of data individually and a model identification tool using decision tree is trained to select proper model for a given operating point. Results showed that MLRMs can be successfully employed in online VSM estimation for large-scale systems.

Despite these fruitful progresses, to the best of our knowledge, most existing learning-based VSM monitoring approaches share two major limitations. First, they assume a global (or large-

area) but relatively simple structure of the predictive model. In other words, these approaches usually assume the predictive model (with determined parameters) could work well under all practical operating conditions. But the model structures (such as polynomial, tree, or practical-sized shallow ANN) are actually not flexible enough to globally describe the underlying pattern, especially when the nonlinear or discrete behaviors of system such as outages and switching controls are considered. Second, they usually train and validate the predictive models purely off-line, with data from limited operating conditions. Therefore, they can neither adapt to the evolving operating condition, nor self-rectify any bad prediction online. [94, 98, 96] proposed some mechanisms to update the predictive models when unseen operating condition is detected. That empowers these methods with some adaptiveness of operating condition, but the "seamless prediction" for unseen conditions is still difficult to be achieved due to the periodical validation strategy or the heavy computational burden for re-training. These unfavorable features of existing methods significantly limit their applicability in the new environment of high variability and uncertainty due to the renewable penetration.

In this work, a novel learning-based VSM online monitoring approach is proposed using local regression and adaptive database. It aims to balance model accuracy and simplicity/transparency, balance global and local patterns, and balance online and off-line trainings. This approach can adapt to the changing condition of system, so it could be more applicable in the system with high renewable penetration.

2.2 Online Control of Long-Term Voltage Stability Margin

Since VSM is a global (entire system related) and implicit (physically unmeasurable) quantity, fast (real-time) local control approaches are usually not valid for VSM maintenance. If near-real-time (usually 5∼15 minutes) is the concerned time framework, optimization or optimal control should be the promising techniques to design the control.

2.2.1 Optimization Based VSM Maintenance

The existing optimization based methods for VSM maintenance can be classified into two categories: indirect methods and direct methods.

2.2.1.1 Indirect methods

Instead of using VSM as the control target, the indirect methods select relevant and manipulatable variables (or state space indices described in 1.1.3 and 2.1.1.1) as the control target, hoping that VSM (or generally the long-term voltage stability) will be effectively enhanced by the control actions as a byproduct.

A typical example for this category is voltage stability enhancement via RPR management [101, 27, 102, 103, 104]. In [27], a dual objective optimization approach to maximize the amount of RPRs and reduce system losses is proposed. Simulation results have shown that the amount of voltage stability margin increased with an increase of reactive power reserves. The approach used a nonlinear optimization framework based on optimal power flow and Benders decomposition to determine the best current operating condition.

The concept of reactive reserve based contingency constrained optimal power flow (RCCOPF) is introduced in [101]. An optimal power flow framework is used to identify the minimal amount of RPRs necessary in order to improve the amount of voltage stability margin for various contingencies and operating conditions. Implementation of the approach shows that the amount of voltage stability margin is improved and that the found system state (power flow solution) corresponds to the minimum effective RPR.

[103, 104] defined the effective RPR directly related to VSM, and maximize them to enhance the voltage stability. [103] also proposed the concept of dynamic RPR that is connected to the short-term voltage stability. [104] considered the control areas and chance constraints. However, in these works, the effective RPR essentially becomes an equivalent long-term voltage stability margin index which cannot be directly monitored and manipulated.

There also exist other indirect methods such as the online voltage/var control (VVC) techniques which consider voltage stability maintenance as one of the objectives [105].

[106] gives a review of many methods in this category from the perspective of voltage constrained reactive power dispatching.

2.2.1.2 Direct methods

The direct methods simply use VSM (the load margin or other margin indices) as the control target (either in the objective or in the constraints). As it will be shown in 3.1.2, VSM is usually a nonlinear, nonconvex, nonsmooth, and even discontinuous function of the operating states and parameters explicit in the system equations (if the function exists). Thus, to integrate VSM into a control optimization problem, we need to either implicitly formulate VSM using (i) bifurcation conditions (like in the direct methods of calculating the collapse point) [107, 108] or (ii) sub-level optimizations (like in the nonlinear optimization methods of calculating the collapse point) [109], or construct an explicit formulation of VSM as a differentiable expression of operating states or parameters. These problems are usually referred to as voltage stability constrained optimal power flow (VSC-OPF). In general, any stability region based method mentioned in 2.1.1.5 or machine learning based method mentioned in 2.1.2 provides a way to obtain such an explicit formulation of VSM.

However, most of these implicity or explicit formulations of VSM are computationally intricate and therefore rarely used in the context of online VSM control. The practical methods in this category are usually based on the linear formulation (approximation) of VSM, i.e., using VSM sensitivity [110, 111]. But calculating VSM sensitivity based on the bifurcation conditions [112] requires a bifurcation point as the initial input, which further requires continuation methods or direct methods to obtain the bifurcation point. In this case, the computational burden may become unaffordable for online application when multiple contingencies are considered.

[113] proposed to address the problem of real time voltage stability through the enhancement of critical RPRs and system VSM. In some sense, this work combines the ideas of both direct methods

and indirect methods – it uses RPRs as the intermediate variables that link control actions to VSM through sensitivities. The method is expected to be used in emergency situations when low amounts of RPRs, VSM or voltage violations are observed. Sensitivities of control actions of critical RPRs are used to determine the optimal amount and location of control. The control problem is formulated as a convex quadratic programming. This approach will be discussed in detail in 4.1.1 and 4.1.2.

2.2.2 Applications of Model Predictive Control in Power System

Model predictive control (MPC) is an advanced method of process control that has been in use in the process industries in chemical plants and oil refineries since the 1980s [114, 115, 116]. In recent years it has been also applied to power system for various purposes such as instability prevention [117, 118], frequency control [119], and voltage/var regulation [120, 121, 122, 123]. [124, 125, 126, 127, 128] applied MPC in voltage stability related controls. Detailed power system model using differential algebraic equation (DAE) including excitor, OLTC and load models, or simplified static models are used to predict the system response to operating condition changes and large disturbances in short-term or long-term time scale. Although the configurations on system model, cost function and available control actions are different in these works, voltage stability constraints are only involved through voltage magnitude constraints. Voltage stability indices such as VSM have never been explicitly involved in these MPC-based methods.

2.2.3 Applications of Demand Response in Voltage Stability Enhancement

The recent report [129] from the IEEE task force on "contribution to bulk system control and stability by distributed energy resources connected at distribution network" summarizes the researches and practices about using demand side controls of active distribution networks for gird stability enhancement. However, the focus of this report is on distributed energy resources (DERs). So far, the only published work we found about using demand response (DR) for VSM enhancement is given by the authors of [111]. Thermostatically controlled loads (TCLs) are engaged in as the control measure. Their capability limits, i.e., discomfort constraints, are converted to instant

feasible ranges of load reduction amount in the bulk level VSM constrained optimization. VSM sensitivity based on [112] is used to locally formulate VSM. Voltage stability security after contingencies are also considered. Our proposed approach follows the idea of using TCLs as an effective VSM control measure, and devotes itself in developing better modeling framework (MPC), TCL capability limits, and VSM formulation techniques.

CHAPTER 3. ADAPTIVE REAL-TIME VSM MONITORING APPROACH VIA LOCAL REGRESSION

3.1 Introduction to VSM Prediction by RPR

This section describes the idea and our previous implementation of predicting VSM using linear regression model of RPRs. The aims of this section are as follows. First, the mathematic essence of the implicit mapping from a few online measurements to the expectation of VSM is discussed to show the theoretical possibility of statistics based approached, and the complexity of this mapping. Second, motivated by our reflection on the physical features of RPR and the statistical data observation, we select RPRs as VSM predictor candidates and use simple model structure. This is inherited by the new proposed approach, and it shows how we integrate our prior domain knowledge into the general machine learning techniques. Third, we briefly introduce our previous work as the base and the reference approach for the new one.

3.1.1 Definitions and Terminology

To facilitate our discussion, first we formally declare some general terms. The *voltage stability margin (VSM)*, denoted by M, refers to the distance between current operating point and the critical point (nose point of PV curve) under certain *hypothetical scenario* measured by the total real power load increment. *Reactive power reserve (RPR)* of a reactive power source j is defined by the difference between the maximal reactive generation Q_j^{max} and the current reactive generation Q_j. Several possible settings of Q_j^{max} will lead to different RPR definitions [100]. In this work we simply use constant Q_j^{max}, which is the standard setting for power flow problem. It is trivial to extend the proposed approaches using some other RPR definitions. $X \in \mathbb{R}^{1\times D}$ refers to the row vector of selected RPRs (as VSM predictor candidates), and D is the number of selected RPRs.

Tracing the PV curve by continuation power flow or other methods gives a series of operating

points, and then yields the corresponding RPRs-VSM, or (X, M) pairs. For a comprehensive long-term *voltage stability assessment (VSA)*, PV curves are traced off-line under various *contingencies* (in this work, force or scheduled outages of generators, transformers and transmission lines are considered), *operating scenarios* (certain system configurations such as generation and load profile, control parameters) and *load increase directions (LIDs)*. This builds up a *database* of (X, M) pairs. The term database will be also used to denote the collection of operating points. Notice that by this setup, a hypothetical scenario, i.e. the way of parameter change, is reduced to a LID.

3.1.2 The Mathematic Essence of VSM Predictive Model

In 3.1.1, we have described how to calculate VSMs under various operating conditions through the procedure of VSA (set parameters of operating condition, then trace PV curve). Our objective is to find the predictive model that explicitly maps online measurements (reflecting the operating condition) to VSM or some approximation of VSM, capturing their relationship that implicitly exists in the procedure of VSA. In this section, we will investigate the abstract equations defining that underlying relationship, i.e., the mathematical equivalent to the VSA procedure. Then we can find that the underlying mapping determined by these equations, i.e., the *learning target* of the predictive model, may generally exist, but usually too complex to be represented by a simple parametric model.

Under reasonable assumptions, the critical point of voltage stability is connected to the saddle-node bifurcation (SNB), which can be described by (ignore transversality conditions)

$$f(x, \lambda, p) = 0 \tag{3.1}$$

$$g(x, \lambda, p) = \det(f_x) = 0 \tag{3.2}$$

f represents the static system equations (usually power flow equations). x is the state variable. λ is the load increment (at a SNB) with respect to certain base value. p is any continuous parameter giving the operating condition, such as the base load profile and the LID. Equation (3.2) (or its variance) is called the bifurcation equation, saying the Jacobian $\partial f/\partial x$ is singular. [112] shows that locally there exists a smooth implicit function h that maps p to λ. If h is the learning target,

then by the universal approximation theorem, it is possible to accurately represent it by a shallow NN.

Unfortunately, (3.1)-(3.2) are not equivalent to the VSA procedure, so h is not our learning target. First, they are valid for all SNBs. But the only SNB making physical sense, which makes λ equal to M obtained by PV curve tracing, is the one connecting stable equilibrium manifold. So, if multiple SNBs exist for fixed p (generally true), extra information is needed to determine the h with physical sense.

Second, (3.1)-(3.2) cannot describe the discrete changes of operating condition in VSA such as bus type switching (e.g. PV→PQ), tap changing, shunt switching, unit start-up/shut-down, and contingencies. These changes are usually regarded as the changes of f and g, so cannot be described universally by a single set of equations. To describe the learning target valid for all operating conditions, we need the equations

$$\bar{f}_1\left(x, M, Y_{bus}\left(p,d\right), p, d\right) = 0 \tag{3.3}$$

$$\bar{f}_2\left(x, M, Y_{bus}\left(p,d\right), p, d\right) = 0 \tag{3.4}$$

$$\bar{g}\left(x, M, Y_{bus}\left(p,d\right), p, d\right) = 0 \tag{3.5}$$

$$\bar{\rho}\left(x, M, Y_{bus}\left(p,d\right), p, d\right) = 0 \tag{3.6}$$

Y_{bus} is the network admittance matrix (notice that most of the discrete changes can be represented by the variation of Y_{bus}). d is any discrete parameter (e.g. on-off states of units, tap positions). $\bar{f}_1$ is the full power balance equations. $\bar{f}_2$ describes the conditional constraints, typically the ones control the bus types (e.g. $\left(Q_i - Q_i^{max}\right) d_k = 0$, where d_k is an element of d such that d_k is nonzero iff Q_i reaches its upper limit). Equation (3.5) represents the bifurcation equation. Equation (3.6) is the condition that restricts the SNB to be the one with physical sense (could be realized by the holomorphic embedding method [130]). Thus, the mapping $\bar{h} : (p, d) \mapsto M$, if uniquely determined by (3.3)-(3.6), explicitly links the parameters to VSM on quite general operating conditions.

However, we do not usually regard $\bar{h}$ as the learning target because p and d are unsuitable to be directly used as the inputs of a predictive model. In particular, (i) $\dim p + \dim d$ (the total

dimension of parameters) is usually very large; (ii) parts of p and d are unobservable from EMS; (iii) parts of p and d are inefficient as VSM predictors. Instead, we select some online measurements

$$y = \mathfrak{O}(x, p, d) \tag{3.7}$$

to be the prediction inputs (RPR, V, and P_{flow} in this work). Since $\dim y \ll \dim p + \dim d$, $\mathfrak{O}^{-1}$ generally does not exist. Thus, $\tilde{h} : y \mapsto M$ is undetermined by (3.3)-(3.7). But if (p, d) has certain joint probability distribution $\mathscr{P}$ for practical operating conditions, then the conditional expectation $\mathbb{E}(p, d \,|\, y)$ can be determined via (3.3)-(3.7). Consequently, there exists $\mathring{h} : y \mapsto \mathbb{E}(M)$ determined by (3.3)-(3.7) and $\mathscr{P}$. The key fact that supports the effectiveness of using $\mathbb{E}(M)$ to predict M is that $\mathscr{P}$ is usually highly sparse and locally low dimensional, such that the deviation of M from its expectation $\mathbb{E}(M)$ is usually at an acceptable level, if proper prediction inputs are selected.

This $\mathring{h}$, if exists, is our learning target. Unfortunately, each of (3.3)-(3.7) or $\mathscr{P}$ is highly nonlinear, discrete and more importantly, full of composite logics. As a result, $\mathring{h}$, which gives the expectation of VSM on any practical operating condition, is very complex and difficult to be accurately learnt by a practical-sized shallow neural network or other parametric models with limited effective capacity.

3.1.3 The Favorable Natures of RPR as VSM Predictor

The possibility and superiority of RPR as VSM predictor are clearly based on its physical natures. First of all, according to [5, 12], for the 2-bus system, given the power factor in normal range, VSM and RPR are restricted on a section of quarter circle, and the system reaches the critical point always when the reactive power generation $q_s = 0.5$ (in p.u. for $S_{base} = E^2/X$), no matter what load power factor it has. This implies that RPR is a good VSM predictor. This analysis and conclusion can be extended to multi-machine system through multi-port Thévenin equivalent [58] (see Appendix A for the derivation in detail).

Beyond this analytic observation, RPRs is endowed with some other favorable features comparing to other online measurements. (i) Explaining VSM by RPRs essentially follows the logic of attributing long-term stability (loadability) to the availability of power sources. This explana-

tion of long-term voltage stability is not strict, but it is logically sound and still an effective and popular conceptual model, especially in industry. (ii) Comparing to the real power reserve, RPR can indicate both real and reactive power load variations. In contrast, real power reserve is not sensitive to the variation of reactive power load through a high X/R ratio network. (iii) Comparing to the voltage, RPR could rarely be blinded (directly regulated) by controllers. In contrast, once a voltage is regulated by switching shunt, FACTS, or OLTC, it cannot see the system changes. (iv) Comparing to the loads, RPR further reflects the transmission loss in addition to the loads. So RPR is more sensitive to load changes especially when the system is close to the saddle-node, given the fact that the sensitivity of loss with respect to load goes to infinity at the saddle-node bifurcation. (v) Comparing to the reactive power generation, RPR carries the extra information of source capability in addition to that of the operating states.

Finally, we can directly observe the correlation between VSM and RPR through the scatterplot of the database (Figure 3.4), which is thoroughly discussed in [100, 131].

All these favorable natures of RPR inspired our attempt to build the VSM predictive model through linear regression over the database of RPRs-VSM pairs.

3.1.4 Building The Database

The first task of establishing the predictive model is to build the database (or the initial database for the proposed approach). For the MLRM approach, it means all information available for extracting the knowledge of the system, whereas for the proposed approach, it just works as the prior information which establishes the "common sense" of the intelligent system. This is to ask how many and which operating points should be included. For our purpose, we only need to include the samples from realistic operating conditions, rather than the whole voltage stability region.

If we have some knowledge of the distribution of the operating condition, there may exist some "optimal" sampling method, such as the variance reduction methods using D-optimality, A-optimality, or G-optimality [132, chap.1], or using active learning techniques [133, chap. 8]. Unfortunately, this is not the case for this study. Even though we know the distribution of operating

condition, it is not proper to be used here, since it cannot reflect the risk of the operating condition, thus the optimality does not fit our values. For example, the severe contingencies in a credible contingency set used for reliability assessment could rarely happen. As a result, if the sampling is based on probability distribution, the operating conditions under severe contingencies can hardly be sampled, thus tend to have larger prediction error. This contradicts to the general goal of stability analysis.

In this work, we assume the prior for the distribution and the underlying model is unavailable. A sufficiently but realistically large range of contingencies, operating scenarios and LIDs should be considered to ensure the generality of the predictive model and incorporate uncertainty. All samples are treated equally. The details of database building will be shown in Table 3.1.

3.1.5 The MLRM Approach

The MLRM approach is described in [100] and [134]. In the off-line phase, the training set is divided into a few groups according to the initial VSM, and for each group, VSM is supposed to be a low degree polynomial (quadratic or cubic in practice) of RPRs plus a homogeneous random error. Then the coefficients of these polynomials can be estimated by ordinary least squares (OLS). These polynomials are represented as the MLRMs. To select the best MLRM for given operating point, a model identification tool (IDTool) based on DT or other classification techniques such as ANN, SVM and KNN (K-nearest neighbor) is also trained off-line using voltage magnitudes and real power flow as inputs, group label as output. In our recent implementation of the method, instead of OLS, LASSO (least absolute shrinkage and selection operator) [135] is used to do the feature selection and linear regression. in the online prediction phase, for given operating point, IDTool selects a MLRM according to the input voltages and power flows, and then the picked MLRM provides the VSM prediction using input RPRs as predictors.

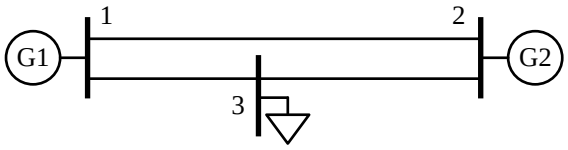


Figure 3.1: Single-line diagram of the 3-bus system. On the base operating condition: $S_L = 15 + \mathrm{j}6$ MVA, $X_{1\text{-}3} = 0.5$ p.u., $X_{2\text{-}3} = 0.6$ p.u., $X_{1\text{-}2} = 1$ p.u., $P_{G2} = 5$ MW, $V_{G1} = 1.05$ p.u., $V_{G2} = 0.95$ p.u. When the load increases, its power factor is fixed, and the increment is dispatched to the two generators, if in-service, according to their reserves.

3.1.6 Limitations of MLRM and Other Methods

As mentioned in the introduction, the MLRM method (and most of other existing ones) has two major limitations. First, MLRM averages the varying local patterns behind a large number of diverse operating conditions. Thus, it sacrifices the accuracy for specific operational condition. Second, MLRMs and DT are trained and validated totally off-line, so the prediction accuracy strongly depends on the quality of off-line database (how many samples are there and how they are distributed), and cannot adapt to the online operating condition.

Figure 3.2 illustrates these problems on the 3-bus system given by Figure 3.1. The RPRs of the two generators are used as VSM predictors. It shows that the RPR-VSM patterns under different contingencies varies a lot, and obviously cannot be summarized by any single quadratic surface in the space. For this simple system, ANN and DT have enough flexibility to separate the curves in the space of Figure 3.2. However, the possible number of curves (contingencies) increases dramatically when the size of system grows, even when we ignore the unpractical or similar ones. That can make a shallow ANN or DT oversized or inaccurate. Further, in the space of RPRs, the projection of the red curve intersects that of the orange one, which implies there are two operating points not distinguishable by the predictive model while their VSMs differ significantly. This is a general issue because in practice, the predictors (here the RPRs) usually cannot uniquely determine the operating condition. Now suppose Line 2-3 is tripped, while this outage, which is distinct, was not included in the contingency set when training the predictive model off-line. Further assume the RPRs are close to the intersection. In this case, any model cannot recognize the true operating

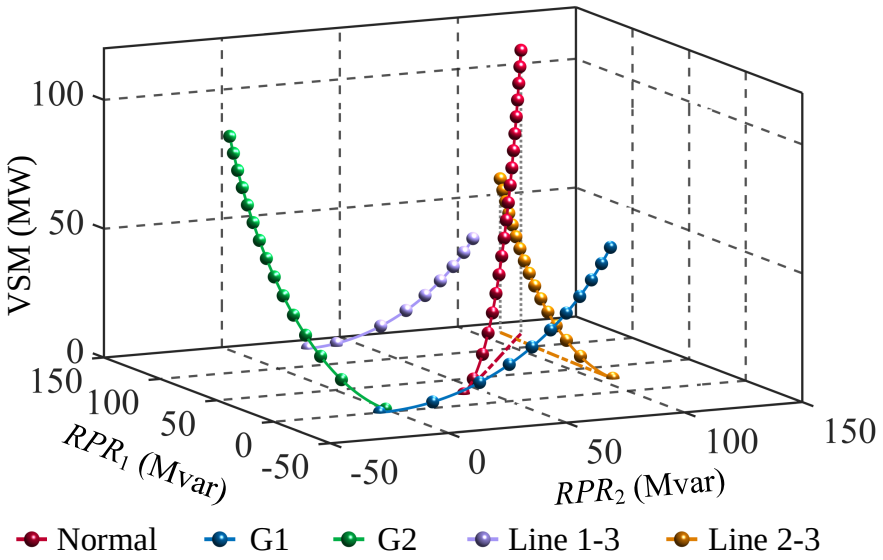


Figure 3.2: RPRs-VSM plots under normal condition and the contingencies. Four contingencies are depicted: G1 outage (blue), G2 outage (green), Line 1-3 outage (purple), and Line 2-3 outage (orange). The vertical projections of the red and the orange curves are drawn on the bottom plane with dashed lines. The operating points are obtained via PV curve tracing under each condition.

condition, and will predict as if the system was operating on the red curve. Such a bad prediction cannot be rectified until Line 2-3 is re-closed.

These limitations could be more remarkable when the renewable penetration is high. In such case, the variation of operating condition, especially the on-off states of conventional generators, could be very large, whereas the off-line training cannot cover all practical combinations of these generator states. In the field test at the control center of a real utility which has more than 50% wind penetration, we observed that MLRMs failed due to half of the conventional generators of selected RPRs are out-of-service.

The recent work [98] admits the locality of underlying model in the topological sense, but the looking-up scheme it uses for topology matching actually gives up learning the topology from data, and consequently cannot directly deal with the unseen ones. Although its model updating procedure empowers the method with adaptability for network topology, but the seamless real-time prediction hardly can be achieved due to the simulation and retraining process for unseen topology. Besides, [98] did not report its nonlinear basis function family, for which we can only find the examples

with cubic polynomial and sinusoidal function restricted within one half period. In this case, the method is essentially an additive polynomial (roughly) model trained by bagging-like approach using heuristic weights. In general, LASSO (as an equivalent of bootstrap for linear model) could perform even better than this kind of (bagging-like) method [136, 137].

3.2 VSM Prediction via Local Regression

To overcome the limitations of existing approaches, we propose to predict VSM via local regression. Local linear regression will be used to train the local predictive model for given operating point. To mitigate the curse of dimensionality, LASSO is used to solve the local regression problem.

3.2.1 Local Regression

The main idea of local regression is to train an *exclusive* model *online* for each given operating point only using the *similar points* in the database. The principle here is to admit and utilize the locality of the underlying pattern which could change a lot due to the discrete behaviors of the system, such as the topology change after contingencies, PV/PQ bus switching when the reactive power generation reaches the limits, and unit starting up or shutting down.

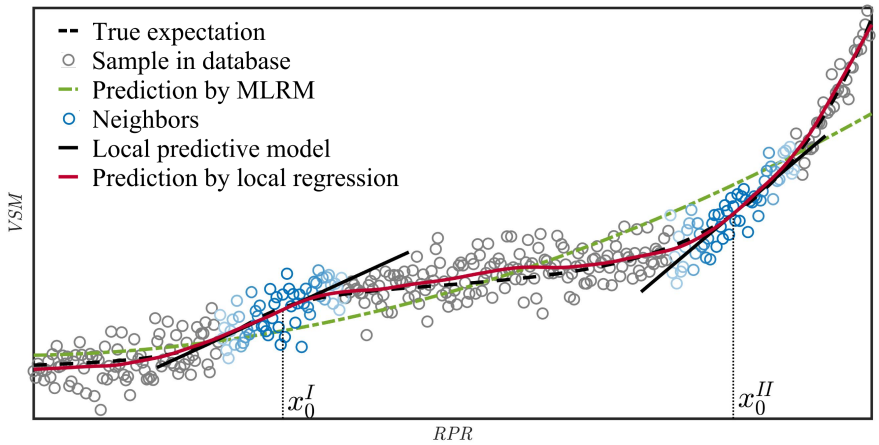


Figure 3.3: Conceptual illustration of local regression.

Figure 3.3 conceptually illustrates this idea. Suppose we have built the database according to

3.1.4, and the data correspond to some MLRM are shown as the circles. The black dashed line shows the true expectation (the best prediction) of VSM given the RPR. So, the target is to reveal it from the circles. A MLRM (green dashed line) can be established by quadratic regression over all database points. However, it deviates from the true expectation in several regions. So instead of using all database points to train the MLRM off-line to predict VSM for all given RPRs within the range, we use only the neighbors (blue circles) of the current given operating point (two instances x_0^I and x_0^{II} are shown) to train the local model (the two black bars for x_0^I and x_0^{II} respectively) online, then use it to predict VSM exclusively for the given operating point. The red solid line shows the predictions of all operating points by local regression, which is closer to the target comparing to MLRM. Two questions follow immediately. One is how to define the neighbors for a given operating point; the other is what regression techniques are used to train the local model.

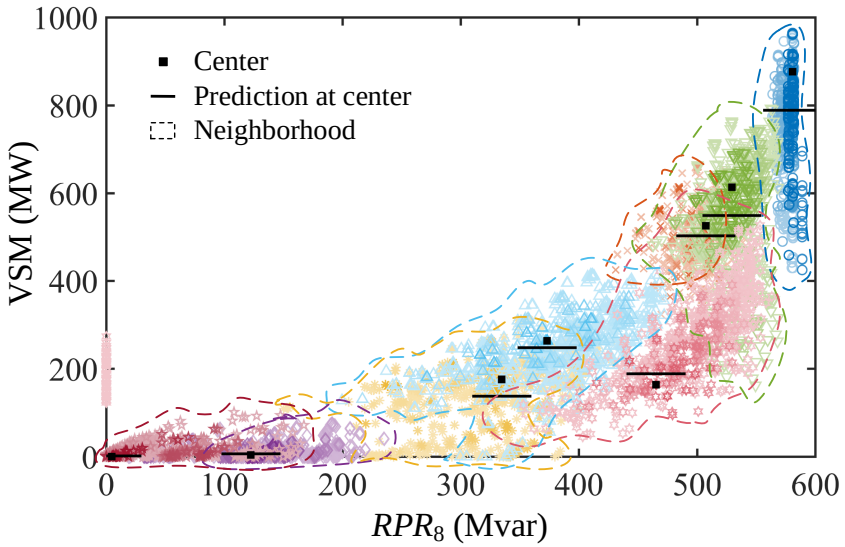


Figure 3.4: K-nearest neighborhoods with tri-cubic kernel for IEEE 30-bus system. RPR_8: RPR of generator at bus 8. $K = 1\%$ of the database size, which is 738 in the example. Each marker (circle, square, triangle, etc.) represents an operating point in the training set. 8 operating points are randomly selected as the centers, denoted by black squares. Their neighbors are annotated by distinct colors and markers, and circled by dashed curves. Darker color indicates larger weight. Local regression gives the predictions at the centers, which are shown as black bars.

A *neighbor* of an operating point is a point in the database which close to it in some physical

sense, such that they are supposed to share the same local pattern. The *neighborhood* of an operating point is the subset of the database which contains all its neighbors. So there are three essential factors to define a neighborhood: space, metric, and size (or boundary). First we need to clarify that in what space the neighborhood is defined. To make the best use of the online measurements, we can define the neighborhood in the space of not only RPRs but also other effective online measurements. As suggested by [134], in this work we include RPRs, voltage magnitudes and real power flows in major transmission lines. Usually, all these measurements are highly collinear. So it is beneficial to reduce this space to a much lower dimension D' by principle component analysis (PCA). Let

$$Z = [X, V, P_{\text{flow}}] \cdot \boldsymbol{C}, \quad Z \in \mathbb{R}^{1 \times D'} \tag{3.8}$$

denote the projection of the online measurements in this reduced space, where V and P_{flow} are the row vectors of monitored voltage magnitudes and real power flows, and $\boldsymbol{C}$ is the projection matrix into the first D' principle components. X, V and P_{flow} are *standardized* to have zero mean and unit standard deviation before PCA. D', the *neighborhood dimension*, is selected to contain 95% of the measurement variance (described in 3.2.2). So the neighborhood is defined in the space of Z, i.e. the *neighborhood space*. Euclidean metric is the natural choice for our purpose. Finally, to control the size of online local regression problem, KNN is applied; i.e. the size of the neighborhood is directly given by its cardinality K. We call it the *neighborhood size*. The value of K is tuned in the validation phase. For a given operating point, we call its projection in the neighborhood space, z_0, the *center* of the neighborhood. On the scatterplot of the database in RPR_8-VSM space, Figure 3.4 shows the neighbors of 8 randomly selected operating points.

Once the neighbors are determined, we can apply certain regression technique to establish the local model. Local linear/polynomial regression is deemed to be the favorable method for our purpose concerning both performance and simplicity. The weights based on tri-cubic kernel are used to emphasize the more relevant samples and smooth the prediction. Next is to determine the degree of polynomial used as the local model. In practice, the degree is up to 3 to prevent over-fitting. In addition, the model with even degree, i.e. 2, usually suffers severer boundary effect.

So, based on the discussion in 3.1.3 and the objective of online application, we choose the degree to be equal to 1 (linear regression). Instead of OLS, LASSO is applied to train the local linear model for three reasons. First, there usually exists high collinearity in the RPRs. Second, when K is relatively small, the design matrix of regression could be not full column rank. In this case, LASSO is used to grantee a unique solution for almost sure. Third, when some RPR is constant within the neighborhood, LASSO can automatically exclude it from predictor candidates. This situation can be caused when Q generation reaches limit, or generator is out-of-service. In all, LASSO is an efficient method that is able to automatically select the local effective predictors while ensure the uniqueness of solution, which is robust for online application.

The local LASSO regression problem can be formulated as

$$\min_{\beta} \frac{1}{2}\left[\boldsymbol{M'} - \boldsymbol{X'}\beta\right]^{\top} \boldsymbol{W}(z_0)\left[\boldsymbol{M'} - \boldsymbol{X'}\beta\right] + \mu\|\beta\|_1. \tag{3.9}$$

Let N be the size of database. $\boldsymbol{W}(z_0) \in \mathbb{R}^{N\times N}$, a diagonal matrix, is the weight matrix corresponding to the current operating point $z_0 \in \mathbb{R}^{D'}$; each of its diagonal element $w_i(z_0)$ is the normalized tri-cubic kernel weight (with KNN window function [90]) corresponding to the ith training point, which is the standard setting for local linear regression [90]. $\boldsymbol{M'} \in \mathbb{R}^N$ is the locally centered VSM vector satisfying

$$\boldsymbol{w}(z_0)\boldsymbol{M'} = 0, \tag{3.10}$$

where $\boldsymbol{w}(z_0) = [w_1(z_0), w_2(z_0), \ldots, w_N(z_0)] \in \mathbb{R}^{1\times N}$. Accordingly, $\boldsymbol{X'} \in \mathbb{R}^{N\times D}$ is the locally normalized design matrix satisfying

$$\boldsymbol{w}(z_0)\,\boldsymbol{X'_k} = 0 \tag{3.11}$$

$$\boldsymbol{X'^T_k}\boldsymbol{X'_k}/N = 1 \tag{3.12}$$

for all k, where $\boldsymbol{X'_k}$ is the kth column of $\boldsymbol{X'}$. β is the coefficient vector of the local model. $\mu \in \mathbb{R}^+$ is the *regularization factor* of LASSO.

There exist quite a few efficient algorithms [137] to solve problem (3.9) in both central or distributed manner. Once $\hat{\beta}_0$, the solution of (3.9), is obtained, we can use the local linear model

to predict VSM at current operating point by

$$\hat{M}_0 = \boldsymbol{w}(z_0)\boldsymbol{M} + x_0'\hat{\beta}_0, \tag{3.13}$$

where $x_0' \in \mathbb{R}^{1\times D}$ is the current predictor normalized by the same operation that maps $\boldsymbol{X}$ to $\boldsymbol{X}'$.

An extra benefit of local LASSO regression comparing to MLRM is that it is usually more robust to the noise and the missing of input RPRs. First, the local model is linear, and the norm of $\hat{\beta}_0$ (a sensitivity of prediction with respect the input noise) is controlled. Particularly, if all predictors have i.i.d. additive random noises with the variation σ^2 (it is not true for the MLRM with high order terms, which is yet more sensitive to the noise), then the prediction variance is

$$\mathrm{var}\left(\hat{M}\right) = \hat{\beta}^\top \Sigma_{X'} \hat{\beta} = \sigma^2 \left\|\hat{\beta}\right\|_2^2, \tag{3.14}$$

where $\Sigma_{X'}$ is the variance-covariance matrix of X'. For the proposed approach, $\hat{\beta}$ is simply the $\hat{\beta}_0$ given by (3.9), and for MLRM, it is a universal coefficient vector and (almost) irrelevant to the current operating point. Notice that the dimension (or the degree of freedom) of the local pattern is usually much lower than the global one, so on average the proposed approach tends to have smaller $\left\|\hat{\beta}\right\|_2$, thus is less sensitive to the noise. Second, if we set a default value for each predictor, namely letting $x'_{0,k} = 0$ by default for all k, then in the case where all RPR inputs are missing without being aware, the local linear predictive model becomes a local constant model by (3.13). This is equivalent to weighted KNN regression and usually works better than MLRM [131]. Notice that VSM is mainly positively correlated to RPRs. Thus, if only a few RPRs are missing, the VSM prediction tends to sit between the non-missing case and the all-missing case. So, the prediction is usually more accurate than the MLRM prediction suffering the same missing. Third, if the missing RPR has been identified, the corresponding column in $\boldsymbol{X}'$ will be deleted, then little accuracy will be lost.

3.2.2 Validation and Parameter Tuning

There are three hyperparameters that must be determined before the local LASSO regression can be done: the neighborhood dimension D', the neighborhood size K, and the regularization

factor μ.

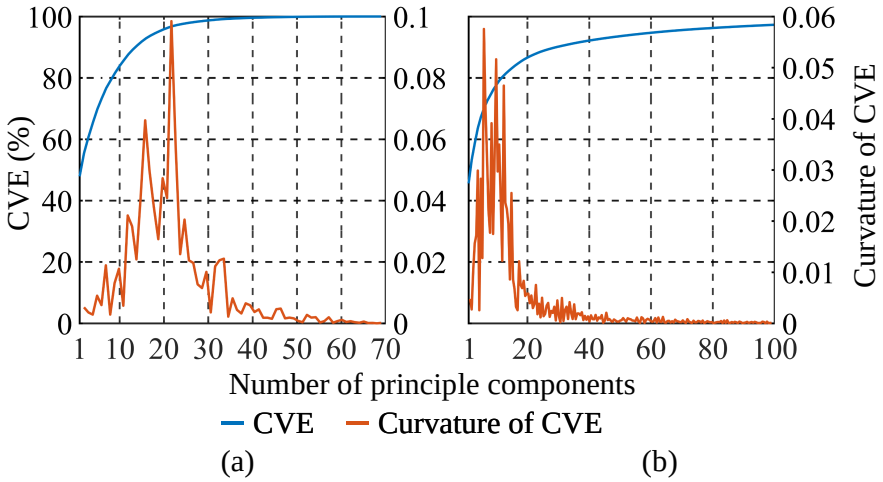


Figure 3.5: Cumulative variance explained of the principle components. (a) is for IEEE30 system; (b) is for REAL system. The AVE curves (blue) share the same scales shown by the left y axes; the curvature curves (red) use different scales shown by the right axes. A point $(10, 82)$ on the CVE curve means the first 10 principle components explained 82% of the measurement variance.

K and μ should be tuned normally in the validation process for the trade-off between bias and variance. For D', nevertheless, the purpose of PCA here is to squeeze out the "useless" dimensions of input measurements that rise from the inherent collinearity of power system state variables [138]. This collinearity is mainly a result of Kirchhoff's laws and the rules of power system operation, so D' is somewhat stable with respect to the variation of operating condition, especially if we set it to be a conservative upper bound of effective dimension. Therefore, the principle for determining D' is that it should be large enough to contain most of the variation of input measurements while computationally practical. In practice, we can select D' according to the plot of *cumulative variance explained (CVE)*, shown in Figure 3.5. CVE can be directly output by most PCA procedures. Usually it increases rapidly at the beginning, then becomes flat (meaning the profit of involving more principle components becomes little). So, a reasonable cut-off number for the retained principle components, i.e. the D', should be on the right side of the "turning-point", which can be mathematically identified by the point with peak curvature. In the work, we set a conservative rule that requires CVE to be greater than 95%. For the test examples shown

in Figure 3.5, the corresponding values of D' are roughly 30 and 60 respectively; both exceed the turning-points. Test results showed that nearly no accuracy improvement could be achieved by further enlarging D'.

μ controls the strength of regularization. For a given operating point z_0, there is a largest μ that permits a non-null model (the one has at least one predictor), denoted by $\mu(z_0)_{\max}$. So, it is unreasonable to assume a universal μ for all operating points, which will force all local models such that $\mu(z_0)_{\max} < \mu$ to be constant. Meanwhile, automatically tuning μ for every local model online is also infeasible because of the computational burden and the relatively limited data (the neighbors). Fortunately, $\mu(z_0)_{\max}$ can be evaluated in advance of regression once the neighbors are given [137]. So, instead of directly tuning μ, we assume there exists a universal ratio

$$\gamma = \mu(z_0)/\mu(z_0)_{\max} \tag{3.15}$$

for all local models, which is called *relative regularization factor*. Thus we can replace μ by $\gamma\mu\,(z_0)_{\max}$ in (3.9), and tune γ instead. This setup enables tuning μ in the off-line phase and locally applying LASSO online. Our testing shows it works.

To understand the construction of γ, suppose there is only one predictor, or in other words, replace $\boldsymbol{X}'$ in the local LASSO problem (3.9) by its kth column $\boldsymbol{X}'_k$. Then it is trivial to derive the solution for (3.9):

$$\hat{\beta}_{0,k} = \begin{cases} \frac{\boldsymbol{X}_k'^{\top}\boldsymbol{W}(z_0)\boldsymbol{M}'-\mu}{\boldsymbol{X}_k'^{\top}\boldsymbol{W}(z_0)\boldsymbol{X}'_k} & \boldsymbol{X}_k'^{\top}\boldsymbol{W}(z_0)\,\boldsymbol{M}' > \mu \\ 0 & \left|\boldsymbol{X}_k'^{\top}\boldsymbol{W}(z_0)\,\boldsymbol{M}'\right| \le \mu \\ \frac{\boldsymbol{X}_k'^{\top}\boldsymbol{W}(z_0)\boldsymbol{M}'+\mu}{\boldsymbol{X}_k'^{\top}\boldsymbol{W}(z_0)\boldsymbol{X}'_k} & \boldsymbol{X}_k'^{\top}\boldsymbol{W}(z_0)\,\boldsymbol{M}' < -\mu \end{cases}, \tag{3.16}$$

which is known as *soft thresholding*. Taking $\boldsymbol{W}(z_0)$ as a Gram matrix, we can define an inner product $\langle\cdot,\cdot\rangle_w$, and a norm $\|\cdot\|_w$. Then the common term in (3.16), i.e., the weighted least square solution, can be written as

$$\tilde{\beta}_{0,k} = \frac{\boldsymbol{X}_k'^{\top}\boldsymbol{W}(z_0)\,\boldsymbol{M}'}{\boldsymbol{X}_k'^{\top}\boldsymbol{W}(z_0)\,\boldsymbol{X}'_k} = \frac{\langle \boldsymbol{X}'_k, \boldsymbol{M}'\rangle_w}{\left\|\boldsymbol{X}'_k\right\|_w^2}. \tag{3.17}$$

In the case where all $\boldsymbol{X}'_k$ are orthogonal to each other by $\langle\cdot,\cdot\rangle_w$, (3.16) completely gives the solution of (3.9). In general, as μ increases, all $\hat{\beta}_{0,k}$ are gradually shrunk to zero. When there is

only one nonzero $\hat{\beta}_{0,k}$ left (suppose (3.9) has a unique solution), (3.16) is also valid. Therefore, there always exists a largest μ that permits a non-null model, which we denote by $\mu(z_0)_{\max}$, satisfying

$$\mu(z_0)_{\max} = \max_{1\leqslant k\leqslant D} \left|\left\langle \boldsymbol{X}'_k, \boldsymbol{M}' \right\rangle_w\right| \tag{3.18}$$

$$= \max_{1\leqslant k\leqslant D} \left|\tilde{\beta}_{0,k}\right| \left\|\boldsymbol{X}'_k\right\|_w^2. \tag{3.19}$$

Notice that $\boldsymbol{X}'$ and $\boldsymbol{M}'$ are centered by (3.11) and (3.10), so $\|\cdot\|_w$ gives their weighted deviations and $\|\cdot\|_w^2$ gives their weighted variances. Thus, $\left|\left\langle \boldsymbol{X}'_k, \boldsymbol{M}' \right\rangle_w\right|$ is the product of the deviation of $\boldsymbol{X}'_k$ and the deviation of $\boldsymbol{M}'$ that can be explained by $\boldsymbol{X}'_k$ (in the sense of projection). Or it can be written as the product of the slope magnitude and the variance of $\boldsymbol{X}'_k$ by (3.19). Therefore, $\mu(z_0)_{\max}$ is a measure of the neighbors' coverage in the product space with respect to X' and M'.

Intuitively, for the local model, we should regularize more when $\mu(z_0)_{\max}$ is large, i.e., when the neighbors are disperse and the regression surface is steep. Thus, in this work, let μ be proportional to $\mu(z_0)_{\max}$ with a universal constant factor $\gamma \in (0, 1] \subset \mathbb{R}$, then we have

$$\hat{\beta}_{0,k} = \begin{cases} \tilde{\beta}_{0,k} - \gamma r_k \tilde{\beta}_{0,k^*} & \left\langle \boldsymbol{X}'_k, \boldsymbol{M}' \right\rangle_w > \mu \\ 0 & \left|\left\langle \boldsymbol{X}'_k, \boldsymbol{M}' \right\rangle_w\right| \leq \mu \\ \tilde{\beta}_{0,k} + \gamma r_k \tilde{\beta}_{0,k^*} & \left\langle \boldsymbol{X}'_k, \boldsymbol{M}' \right\rangle_w < -\mu \end{cases}, \tag{3.20}$$

for all k in the case where all $\boldsymbol{X}'_k$ are orthogonal to each other, or for $k = k^*$ in general, where k^* corresponds to the last nonzero $\hat{\beta}_{0,k}$ as μ increasing, and $r_k = \left[\left\|\boldsymbol{X}'_{k^*}\right\|_w / \left\|\boldsymbol{X}'_k\right\|_w\right]^2$. Equation (3.20) shows how γ controls the strength of regularization.

In this way, we only need to tune the single hyperparameter γ in the validation. By contrast, if we assume a universal μ, then for all z_0 such that $\mu(z_0)_{\max} < \mu$, $\hat{\beta}_0$ is forced to be all-zero, which is unreasonable; if we tune μ online for each z_0, the computational burden is high and the data are scarce.

Finally, from the Bayesian perspective, the heuristic of γ indeed adopts the Laplacian prior density

$$\mathscr{P}_\beta(\beta) = \left[\frac{\gamma\mu(z_0)_{\max}}{2}\right]^D e^{-\gamma\mu(z_0)_{\max}\|\beta\|_1}. \tag{3.21}$$

Strictly, this is not a "prior" since $\mu(z_0)_{\max}$ depends on data. But it expenses only at most D degrees of freedom, so the effect can be ignored when $D \ll K$. Let $\alpha = \mu(z_0)_{\max}\beta$. Equation (3.21) implies, for various z_0, $\mathscr{P}_\alpha(\alpha)$ is invariant:

$$\mathscr{P}_\alpha(\alpha) \equiv \left[\frac{\gamma}{2}\right]^D e^{-\gamma\|\alpha\|_1}. \tag{3.22}$$

Then we can show the irrationality of our heuristic by the dimensional analysis. For any quantity $\mathcal{X}$, let $[\![\mathcal{X}]\!]$ denote the unit of $\mathcal{X}$. For weighted least square problem, it is normal to regard the weight matrix as the inverse of the variance-covariance matrix $\Sigma_{X'}$ whose unit is $[\![\boldsymbol{M'}]\!]^{-2}$. Thus, $[\![\alpha]\!] = [\![\mu(z_0)_{\max}\beta]\!] = [\![\boldsymbol{X'}]\!][\![\boldsymbol{M'}]\!]^{-2}[\![\boldsymbol{M'}]\!]\cdot[\![\boldsymbol{M'}]\!][\![\boldsymbol{X'}]\!]^{-1} = 1$. That seems reasonable to let a dimensionless quantity have an invariant distribution. Also notice that γ is dimensionless due to (3.15), which is also logically sound for a constant parameter.

For jointly tuning K and γ, the widely used cross-validation suffers from over-fitting for our study because the number of considered contingencies, operating scenarios and LIDs are much less than the size of the database, so all of them are quite likely be traversed by each fold of training set. Thus, we generate a separate validation set, which includes some exclusive contingencies, operating scenarios and LIDs. Thus, the tuning in validation is to solve the optimization problem

$$\begin{aligned} &\min_{K,\gamma} \quad \sqrt{\tfrac{1}{N^v}\sum_{j=1}^{N^v}\left[M_j'^v - x_j'^v\hat{\beta}_j(K,\gamma)\right]^2} \\ &\text{s.t.} \quad \text{for all } j: \\ &\hat{\beta}_j(K,\gamma) = \underset{\beta}{\operatorname{argmin}} \left\{\tfrac{1}{2}\left[\boldsymbol{M}'^t - \boldsymbol{X}'^t\beta\right]^\top \boldsymbol{W}^t\left(z_j^v\right)\left[\boldsymbol{M}'^t - \boldsymbol{X}'^t\beta\right] + \gamma\mu\left(z_j^v\right)_{\max}\|\beta\|_1\right\} \end{aligned} \tag{3.23}$$

where the superscript "v" indicates the quantities of validation set, while "t" indicates the quantities of training set. The objective function is the root-mean-square error over the validation set, given $\hat{\beta}_j(K,\gamma)$ for all j, and N^v is the size of the validation set. For each j, $\hat{\beta}_j(K,\gamma)$ is the solution of (3.9), given a point in validation set, z_j^v, as the current operating point. Notice that (3.9) is solved using neighbors in the training set, for fixed K and γ. Since there are only two continuous decision variables in this problem and the searching space is convex, it is trivial to solve the problem by

general optimization solver, or simply by manually searching (demonstrated in Figure 3.9). For the test examples, solutions with acceptable precision are obtained within 10 iterations.

3.3 Adaptive Database

The initial database trained off-line cannot always densely cover the current operating region, especially when high renewable penetration and system expansion are considered. So, an effective database should be updated automatically to adapt to the system changes.

3.3.1 Criteria for Triggering Database Updating

To control the cost of computation and data storage, it is unnecessary to augment the database unceasingly. So, the first task to design the database updating scheme is, naturally, to detect the influencing system changes that really need an updating, i.e. the ones that possibly degrade the VSM prediction accuracy.

A straightforward idea is based on the prediction interval. If the prediction interval of each prediction can be provided, the influencing system change is identified if the prediction interval exceeds some threshold.

The coarsest prediction of prediction interval may be the one based on the empirical error distribution of the whole database (described in 3.5.3) or its bootstrap version, which is almost fixed and does not fit our goal. On the contrary, a very fine prediction, which does not assume any homogeneity of bias or variance over the predictor space, requires an extra nonparametric learner of similar complicity in addition to the one for predicting VSM, which significantly increases computational burden and sample complicity. We propose to use the compromised strategy: assuming the approximated homogeneity of bias and variance for each local model. Under this assumption, the *weighted root-mean-square error over the neighbors (locRMSE)*, defined as

$$locRMSE_0 = \left[\left[\boldsymbol{M}' - \boldsymbol{X}'\hat{\beta}_0\right]^\top \boldsymbol{W}(z_0) \left[\boldsymbol{M}' - \boldsymbol{X}'\hat{\beta}_0\right]\right]^{1/2} \tag{3.24}$$

is a natural estimate of the absolute prediction error within the neighborhood of z_0. Then it is reasonable to estimate the prediction interval through the error empirical distribution conditional

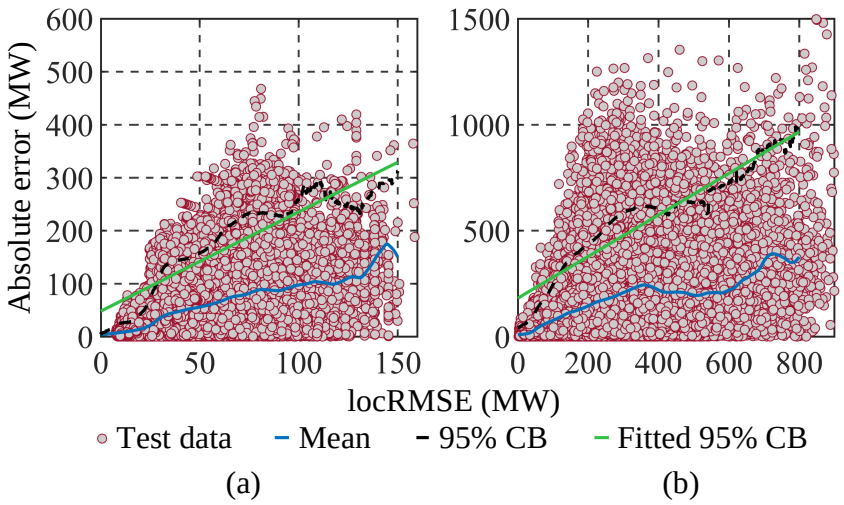


Figure 3.6: Absolute prediction error vs. locRMSE. (a) is for IEEE30 system; (b) is for REAL system. Each circle gives the locRMSE and absolute prediction error corresponding to one operating point. The blue curve gives the mean absolute error conditional to locRMSE. The black dashed curve gives the empirical 95% confident upper bound conditional to locRMSE. The green line is the linear fitting result of the black curve, which gives the affine mapping f from locRMSE to the prediction upper bound.

to locRMSE.

This empirical distribution is almost ready-made. During each local regression process, LASSO solver can output locRMSE without imposing extra computation. Thus we can obtain the scatterplot of absolute prediction error (denoted by $|e|$) vs. locRMSE (Figure 3.6) for free during the off-line validation or test. Then the empirical mean (blue curve) and confidence upper bound (black dashed curve) conditional to locRMSE follow immediately (detailed in Appendix B). Figure 3.6 validates our homogeneity assumption (the blue curves are nearly linear with unit slope), and also implies that (i) although locRMSE alone is not enough to explain all variance of absolute error, (ii) it is a significant linear predictor for the mean and the confidence upper bound of the absolute error. Suppose the linear fitting of the empirical 95% confidence upper bound (the black dashed curve) gives an affine function $f : locRMSE \mapsto |e|$ (the green line in Figure 3.6), then the *VSM prediction interval* at z_0 is

$$PI_0 = \left[\hat{M}_0 - f(locRMSE_0)\,, \hat{M}_0 + f(locRMSE_0)\right] \tag{3.25}$$

Notice that it can be monitored in real time, synchronized to the VSM prediction. Once it exceeds certain threshold, the database should be updated. Since f is affine, it is also convenient to directly preset the threshold for locRMSE (denoted by s_1), rather than PI, and monitor locRMSE as an index of prediction confidence. So, our first criterion (C1) to trigger the updating is: if $locRMSE_0 > s_1$, start the updating.

Besides, some large system changes that are scheduled or predictable (such as line switching, maintenance, start-up/shut-down of units, and some large change of dispatching due to the wind variation) can be used to trigger the update. Operators can list them as the *set of possible influential events*, denoted by G. If some events in G will happen in near future according to any schedules or forecasting, what-if scenarios can be generated in OLPVE. Then the database updating can be accomplished in advance of the actual system changes, achieving the "seamless adaptation". So our second criterion (C2) to trigger the updating is: if $\{\text{near-future events}\} \cap G \neq \varnothing$, start the updating. We limit the discussion on G because it can be suggested by experience or off-line stability studies (routines for many operators).

Thus, the database augmenting is triggered correctively by (C1) and predictively by (C2), or manually triggered whenever the operators think it is necessary.

3.3.2 Database Augmenting

When (C1) and (C2) indicate some unfavorable system change, a warning will be sent to operators, and the database should be locally updated to improve the prediction. Particularly, more operating points relevant to the current one should be added to the database. Lots of existing EMSs has the capability of tracing PV curves for a bunch of contingencies and LIDs in near-real time (e.g. 1$\sim$10 minutes). we propose to utilize such an online PV curve tracing engine (OLPVE) in EMS to generates the operating points relevant to the current one. They can be obtained by adding small perturbations with empirical distributions, or high-probability/critical contingencies to the current operating point, then tracing PV curves starting from them, along randomly selected LIDs (based on energy and load forecasting). Then the new generated data are added in to the

database. We refer to this process as *OLPVE augmenting*. Assume that (i) on normal condition, the operating states do not change too fast with respect to the near-real-time scale (usually a mild assumption), and (ii) if some contingency occurs, it is within the high-probability/critical contingencies of OLPVE. Then, the recent generated operating points by OLPVE are quite likely to be close to the current one. Thus, the KNN rule will preferentially choose these points as neighbors, and consequently decrease the prediction error. In other words, OLPVE augmenting makes the neighborhood tight and dense, which enhances the local training data and helps to alleviate the curse of dimensionality.

The number of considered perturbations, contingencies and LIDs for OLPVE augmenting, depend on hardware capability and should be based on practice. The neighborhood size K is a reasonable default setting for number of new operating points, which allows all new data to be included in the neighborhood. Notice that PV curves can be traced in parallel, so this process can be speeded up if more computational units are involved. In section 3.5, significant prediction improvement can be achieved even without considering any perturbation and contingency.

During the OLPVE augmenting, operators should keep skeptical to the prediction and pay attention to locRMSE, which still provides some clue of prediction error in this period. OLPVE augmenting can be continued for a few rounds, and terminated whenever (C1) and (C2) are eliminated.

Due to the OLPVE augmenting, the size of database may grow to unacceptable level. In this situation, some forgetting schemes [139] can be used to screen out ineffective data from the database. We limit our discussion about the forgetting scheme due to the limited space.

3.4 Overall Framework

Figure 3.7 shows the overall framework of the proposed approach. We have three guidelines for this framework design. First, integrate off-line and online information (common sense + situational knowledge) to get the adaptive prediction; second, provide real-time prediction, as well as the real-time inference (namely the prediction interval) simultaneously; third, reduce the dimension of

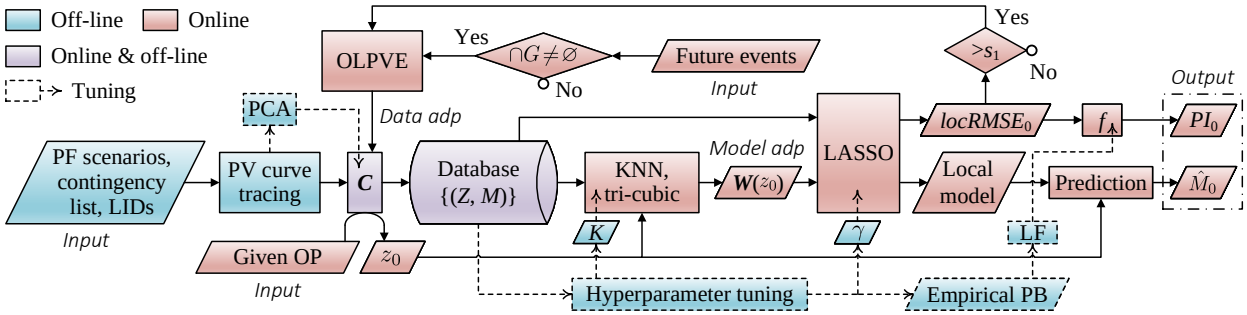


Figure 3.7: The flowchart of the proposed approach. adp: adaptation; $\boldsymbol{C}$: the projection matrix of PCA; f: the affine mapping from locRMSE to the fitted PB; G: the set of possible influential events; LF: linear fitting; $locRMSE_0$: local RMSE over the neighbors of z_0; $\hat{M}_0$: the VSM prediction of current OP; OP: operating point; PB: prediction bound; PF: power flow; PI_0: the prediction interval of $\hat{M}_0$; s_1: the preset threshold of $locRMSE_0$; $\boldsymbol{W}(z_0)$: the weight matrix corresponding to z_0; z_0: the projection of given OP in the neighborhood space. No action is needed for "No" at the two decision blocks (diamond).

feature space to mitigate the curse of dimensionality and enhance the scalability.

In particular, the main procedures of the proposed approach can be summarized as follows. For the off-line phase:

1. An initial database of operating points (including the corresponding VSMs) is built by the VSA over a number of suggested and randomly selected contingencies, scenarios and LIDs (discussed in 3.1.4 and 3.5.2).

2. PCA (including tuning D') is conducted to get the projection $\boldsymbol{C}$ that reduces the states of an operating point to a D'-dimensional vector Z, leading to the database of (Z, M) pairs (discussed in 3.2.1 and 3.2.2).

3. The database is divided into three parts, i.e. training set, validation set and test set. Together with the training set, the validation set is used to tune the hyperparameters K and γ (discussed in 3.2.2), as well as determine the affine mapping f (discussed in 3.3.1), whereas the test set is used to evaluate the performance (generalization error) of the approach (shown

in 3.5.3). Once the validation and the test are passed, the entire database is stored in memory and ready for the online procedures.

For the online phase:

4. Given the current operating point from EMS, RPRs, voltage magnitudes, and power flows are converted to a vector z_0 by the projection $\boldsymbol{C}$ (via (3.8)).

5. The neighbors of z_0 in the current database (initial generated + online updated data) are identified by KNN and weighted by tri-cubic kernel (discussed in 3.2.1).

6. The local predictive model is trained by LASSO, using the neighbors and the corresponding weights (embodies the adaptivity of model; discussed in 3.2.1).

7. Combined z_0 (or x_0) as inputs, the local model gives the prediction of the current VSM, i.e. $\hat{M}_0$ (via (3.13)).

8. Meanwhile, LASSO also outputs $locRMSE_0$, which is then converted to the prediction interval of the current prediction, i.e. PI_0, by the affine mapping f (via (3.25)).

The online procedues, i.e. 4)-8), will be continuously conducted for each operating point given by EMS. Meanwhile:

9. Check the database updating criteria (C1) and (C2). (i) Does locRMSE exceed the required threshold s_1? (ii) Will there be any possible influential events in the near future? (Described in 3.3.1.)

10. Once the updating criteria, either correctively or predictively, are triggered, the OLPVE generates relevant operating points online and add the corresponding new (Z, M) pairs into the database. This process integrates the online information into the database and makes it adaptive to the current operating condition (embodies adaptivity of data source; described in 3.3.2).

Finally, it is worth noting that there is no permanent distinction between the off-line (old) and the online (timely) data. As time goes by, the online data generated by OLPVE becomes "off-line". In other words, the situational knowledge is absorbed by the knowledge pool (database), and becomes part of the "common sense" of the intelligent system.

Several off-line tuning processes are involved (described in 3.2.2 and 3.3.1). The neighborhood dimension D' is selected according to the CVE plots over the whole initial database. The neighborhood size K, and the relative regularization factor γ are then tuned using the training and validation data in a closed-loop optimization process. In the last iteration of this optimization, where the best K and γ are achieved, LASSO also produces the locRMSE-error scatterplot (e.g. Figure 3.6) over the validation set. Then the affine mapping f, can be obtained by simply the linear fitting on the empirical prediction bound.

Based on the three guidelines mentioned at the beginning of this section, the proposed framework extends the technology of real-time learning-based VSM prediction in terms of:

1. combining the off-line and the online data, via the adaptive regression algorithm (local linear regression) and the adaptive data set, to provide timely VSM prediction on the changing operating condition;

2. providing the time-varying estimation of the prediction interval along with the VSM prediction, so the operators can get the sense that how they can trust the current VSM prediction and where the true value of VSM could locate, then the closed-loop corrective adaptation can be established (bad prediction can be automatically rectified);

3. combining local linear regression and LASSO via the relative regularization factor, so as to achieve the sufficient scalability for large-scale power systems.

They are deemed as the major contributions of this work.

3.5 Examples

First let us make an argument on the principle for comparing the proposed approach to others. The performance of a learning based approach in this field significantly depends on the system setting, sampling and algorithm implementation. So, it is imprudent to compare the accuracy numerically on arbitrary examples and draw some general conclusion, unless we can reproduce others' implementations according to the published information and build a uniform base for comparison. Therefore, we decide to compare the performances of the proposed approach only to the reference one, i.e. the MLRM approach we published previously in [100]. Some qualitative comparisons to other approaches can be found in [100, 98].

3.5.1 The 3-bus System

As a preview of the effect of the proposed approach, Figure 3.8 visualizes the prediction results on the 3-bus system given by Figure 3.1 and Figure 3.2. The training set only contains the operating points under normal condition, G1 outage, and Line 1-3 outage. The validation set also includes the points under G2 outage, in addition to the former three conditions. For the proposed approach, all 2 RPRs, 3 voltage magnitudes and 3 active power flows are used as input measurements; $D' = 4$, $K = 116$, $\gamma = 0.731$, where K and γ are tuned in Figure 3.9.

Figure 3.8(a) shows the limitation of a single quadratic predictive model. For the operating conditions in the training set, it gives good prediction on normal condition (the red cubes are very close to the pink spheres), but under G1 (blue) and Line 1-3 outages (purple), the error is large when the operating point is close to the critical point. It gives very bad predictions for the two unforeseen operating conditions (green and orange). Following the approach in [134], two MLRMs are used in Figure 3.8(b). The predictions under G2 outage (green) is improved, but it still fails to predict under Line 2-3 outage. Moreover, as mentioned in 3.1.6, this kind of improvement is unscalable due to the complicity of a real large system, and the bad predictions by off-line MLRMs can never be self-rectified online. In Figure 3.8(b), the local LASSO regression achieves some improvement comparing to Figure 3.8(b), especially for Line 2-3 outage on stressed conditions. When applying

the proposed approach online, the unforeseen contingency (Line 2-3 outage), if occurs, will trigger database updating, then the prediction becomes also quite accurate under this contingency (shown in Figure 3.8(d)). Besides, the undistinguishable issue at the intersection point mentioned in 3.1.6 is fixed, due to the extra information contained in V and P_{flow} of the neighborhood space. The performance of the proposed approach will be evaluated statistically on other two practical systems in the following section.

3.5.2 Database and Parameters

Two example systems are used to test the proposed approach. The first one is IEEE 30-bus system, which has 6 generators. The second one, denoted by REAL, is an equivalent of a large-scale system with emphasis on the footprint of a real utility. It has more than 62,000 buses ranging from distribution level to 750 kV, 1000 generators in total, and more than 600 buses, 100 generators within the utility's footprint. For the second test system, we play as the control center of the utility and monitor the VSM for its operational footprint.

A summary of the database is shown in Table 3.1. The neighborhood dimension D' are 30 and 60 for IEEE30 and REAL system respectively (tuned in Figure 3.5). The hyperparameters are tuned using the validation set. For IEEE30 system, $K = 3321$, $\gamma = 0.077$; for REAL system, $K = 1000$, $\gamma = 0.928$. The γ for REAL system is close to 1 (the local model is close to be constant) because the dimension of X and Z are relatively high whereas the initial validation and test set are intentionally built using a large proportion (half) of unforeseen operating conditions. Equipped with the updating mechanism, the online database will tend to contain more realistic and relevant operating points. So, in practice, operators can re-tune the hyperparameters for a few rounds after some periods of online application, using the online database. Then, foreseeably, γ will decrease and more RPR terms will be involved in the local model.

Table 3.1: Summary of The Database

System	Num_Con				Num _LID	Size	Num _RPR	Num _Mes
	N-1	N-2	N-3	Sev				
IEEE30	43	259	80	0	40	206621	6	71
REAL	339	200	200	51	40	415139	85	2074

Num_con: counts of contingencies. Sev: severe N-k contingencies suggested by utility. Num_LID: counts of LIDs. Size: counts of operating points in the database. Num_RPR: counts of monitored RPRs, including all nonconstant RPRs within study areas. Num_Mes: counts of monitored online measurements, including RPRs, voltage magnitudes and real power flows.

N-1 contingencies include outages of all in-area generators, transformers, and transmission lines (above 161 kV). N-2 and N-3 contingencies are random combinations of N-1 contingencies. LIDs are randomly sampled from Gaussian distribution with the base load profile as the expectation. The database is divided into the training set (50%), validation set (25%) and the test set (25%). 25% of the N-2, N-3 contingencies, LIDs are preserved for validation and test set respectively. E.g. 10 LIDs in validation set are never used in training set; 10 LIDs in test set are never used in training and validation set. So, the contingencies, operating scenarios and LIDs in validation sets and test sets are roughly half known and half unforeseen.

PSS/E is used for VSA (PV curve tracing). OLTCs and HVDC taps, switched shunts work in auto-mode during the VSA. For REAL system, ±250 MW perturbations on redispatching of generations and loads, and ±0.02 p.u. perturbations on generator scheduled voltage are applied to the base point of each PV curve.

3.5.3 Static Prediction Accuracy

We refer to the prediction accuracy without the effect of database updating as *static accuracy*, which implies that all predictions are treated independently. It is assessed using the tuned hyper-parameters and the separate test set. The results are shown in Table 3.2 and Figure 3.10.

We follow the procedures in [100] and preset the RMSE targets as 10% for IEEE30 and 5% for REAL. To simplify the comparison, we exclude the error imposed by IDTool. So, the prediction errors of MLRM approach in TABLE 2 are actually the lower bounds. From Table 3.2 we can see the proposed approach achieves better static accuracy comparing to the MLRM approach (column 3 and 5). Particularly, 21.84% error reduction for IEEE30, and 16.6% for REAL (both using LASSO). More improvement is achieved on IEEE30 mainly because MLRM is underfitting

Table 3.2: Static Prediction Error in RMSE (Unit: $\text{VSM}_{\text{initial}}\%$).

System	MLRM		Local regression		
	OLS	LASSO	OLS	LASSO	Structured
IEEE30	9.14	9.11	7.47	7.12	7.24
REAL	202.37	5.18	15.04	4.38	4.32

RMSE: *root-mean-square error* over the test set, i.e. the generalization error. $\text{VSM}_{\text{initial}}$: base point VSM on normal condition. For IEEE30, $\text{VSM}_{\text{initial}} = 964.06$ MW; for REAL, $\text{VSM}_{\text{initial}} = 4912.5$ MW. Structured: the same as the local LASSO regression except that RPRs are excluded from the neighborhood space (no X in (3.8)).

MLRM use quadratic models, and for REAL, it is a quadratic function of the first 30 principle components of the 85 RPRs. Errors of MLRM excludes the error imposed by IDTool.

OLS uses all predictor candidates without feature selection, and the local OLS regression tunes K in validation phase. For REAL, the best K is almost 100% of training data.

for it (only 6 RPRs). The actual accuracy improvement could be significantly larger considering the effect of database updating (see 3.5.4).

In the last column, Table 3.2 also shows the result for structured local model (varying coefficient model), which is deemed to be more suitable for high dimensional problem than typical local regression [90]. It can be obtained by simply removing RPRs from the derivation of Z (see (3.8)). The test results show that the local LASSO regression in our problems performs roughly the same as the structured local LASSO regression, possibly due to the correlation between RPRs and other online measurements. That implies the general weakness of local regression is not prominent in our study.

One of the underlying motivations of this work is that the static RMSE, as an average, cannot tell the whole story. As in many other literature, the static RMSE of the test example seems quite acceptable for industry application. Here even MLRM for REAL system has only 5.18% error. However, if we look at the error histograms shown in Fig. 4, there are long tails beyond the RMSE for both approaches. In fact, data generated purely off-line cannot always cover the real online

operating conditions, and extremely bad prediction can never be completely eradicated. We find that local regression also wins in this perspective because the more concentrated and symmetrical histogram implies the evaluation metric (RMSE) of the approach is more trustable. But more importantly, we design the adaptive database to rectify the extreme predictions. Let's show this effect through a simulation of online adaption.

3.5.4 Online Adaption

Suppose the prediction is made every 10 seconds. Each round, OLPVE can trace PV curves starting from the current operating point, along 20 randomly selected LIDs within 100 seconds. Two simulations of online adaptive predictions are presented in Figure 3.11, which shows the VSMs (true and predicted) and locRMSE over the time.

Suppose the operator requires the prediction error to be less than 400 MW (around two times of the RMSE), which is around 8% of the margin under normal condition. According to the blue line in Figure 3.24(b), we use $s_1 = f^{-1}(400) \approx 160$ MW as the threshold of (C1). The system was initially operating on normal condition. Loads were slowly increasing with uncertainty. A critical transmission line tripped at t = 50 (denoted by E1). Then two major generators with 1600 MW generation in total tripped at t = 210 (denoted by E2). The tripped line then was reclosed at t = 370 (denoted by E3). After E1, although the contingency, an N-1, was included in the offline contingency set, locRMSE exceeded $\text{locRMSE}_{\text{th}}$, so updating was triggered. 20 PV curves from the operating point at that moment were traced during the next 100 seconds, and 1000 generated operating points are added in the database at t = 150 (A1). OLPVE augmenting is also trigged after E2 and E3 according to (C1) or (C2). In Figure 3.11(a), the proposed approach always beats MLRM, and becomes quite accurate after OLPVE augmenting. Over the whole simulation, the RMSE of proposed approach reduced by 46% with respect to MLRM. The yellow area is the 95% prediction band given by (3.25), which covers the true VSM all through the simulation. Finally, we find MLRM could make large error when major RPR predictors become unavailable due to generator outages (intentional or forced), this can observe in E2. The proposed approach will be

less affected, because LASSO tends to exclude those RPRs from the local model due to the low variation within the neighborhood (mostly fixed at zero).

During the short period between the updating is triggered and it is finished, it is possible that local regression is less accurate than MLRM. Figure 3.11(b) shows one extreme case. A different transmission line tripped at t = 50 from normal operating condition (denoted by E). Then the prediction error of proposed approach significantly increased and even doubled the error of MLRM. This situation is rare according to Fig.4, and it is captured by locRMSE and self-rectified by one round of OLPVE augmenting after t = 150 (denoted by A).

In summary, mostly, (C1) and (C2) can effectively capture the unfavorable system changes and trigger the database updating that ensures the prediction accuracy with certain confidence. Equipped with this dynamic database, the proposed approach can adapt to the changing condition of system and achieves better prediction accuracy comparing to the static result shown in Table 3.2. During the OLPVE augmenting, the prediction becomes less trustable, but the prediction is seamless and locRMSE can still give some clue about the prediction error.

3.5.5 Computational and Data Storage Considerations

Analysis of computational and data storage complexity is necessary for a scalable online method. Here a personal computer with 4 cores, 2.9 GHz CPU and 32 GB memory is used to estimate the time usages.

(Initial) database building, validation (including hyperparameter tuning), testing and MLRM training are accomplished off-line. Python and Matlab scripts are written to automate this process, and normally it takes 1∼2 days.

The major concern is in the online phase. First, we need to consider the RAM usage of the database. If we use double precision to store a number, the RAM usage is roughly $(D + D') \times N \times 8$ bytes, which is 459.25 MB for the REAL system. This is affordable for common EMS server. With the help of LASSO and the adaptive database, this usage increases sublinearly with the system size. Second, the time usage for online prediction, mainly the neighbor identification and local

regression, is the top concern for the whole approach. For REAL system, searching the K-nearest neighbors using exhaustive NNS algorithm of Matlab takes 0.19 seconds on average, and each LASSO regression takes less than 0.1 seconds. In general, searching the K-nearest neighbors from $\boldsymbol{Z} \in \mathbb{R}^{N \times D'}$ could be achieved with time complexity at most $O(D'KN \log N)$, and effective LASSO algorithm like LAR can achieve the same order of computation as a least squares fit, which has time complexity $O\left(ND^2\right)$ for the system has D RPR predictors. The actual time usage of LASSO also depends on the degree of freedom (DOF) of the local model, which is usually low. In practice, the time usage is dominated by D. For a system with thousands of generators, pre-selection of RPRs may be necessary to limit the value of D. Numerous existing feature selection tools can be used to achieve this goal. In summary, practically each online prediction for a real system can be achieved within 1 second, neglecting the data collection and communication time. Thus, considering the space and time complexity, the proposed approach is competent at online application for the system even much larger than REAL system, from efficiency point of view.

3.6 Conclusion

An online VSM monitoring approaching using local regression with adaptive database is proposed in this work. A database that incorporates a large range of uncertainties is initialized off-line through the same procedure for MLRM. For a given operating point online, it first identifies its K-nearest neighbors in the database, then only use these neighbors to train a local predictive model exclusively for the given operating point. LASSO is used to train the local model. The database is designed to be self-updating. Two criteria based on the prediction interval (or locRMSE) and the set of possible influential events, as the updating triggering signals, are proposed to identify the risky operating condition on which the VSM prediction is unreliable. Then the online PV curve tracing engine in EMS is supposed to generate operating points relevant to the current one and add them into the database. Thus, the prediction error can be maintained within given threshold with certain confidence. The prediction interval is provided simultaneously with the VSM prediction, which gives the clue about how the prediction can be trusted and where the true value could be.

Two examples are presented to demonstrate the accuracy, efficiency, adaptability and the scalability of the proposed approach.

Generally, local regression suffers the curse of dimensionality. It achieves favorable performance in our study for the following reasons. First, for a large system, the underlying pattern has local dimension much lower than the number of predictors. LASSO is applied to utilize this feature and reduces the dimension of the predictor space towards the local dimension. Second, the adaptive database keeps filling the neighborhood with new generated data and makes it tight and dense, which mitigates the curse. Finally, according to test result, our predictive model can be regarded as structured local model, which is more suitable for high dimensional problem.

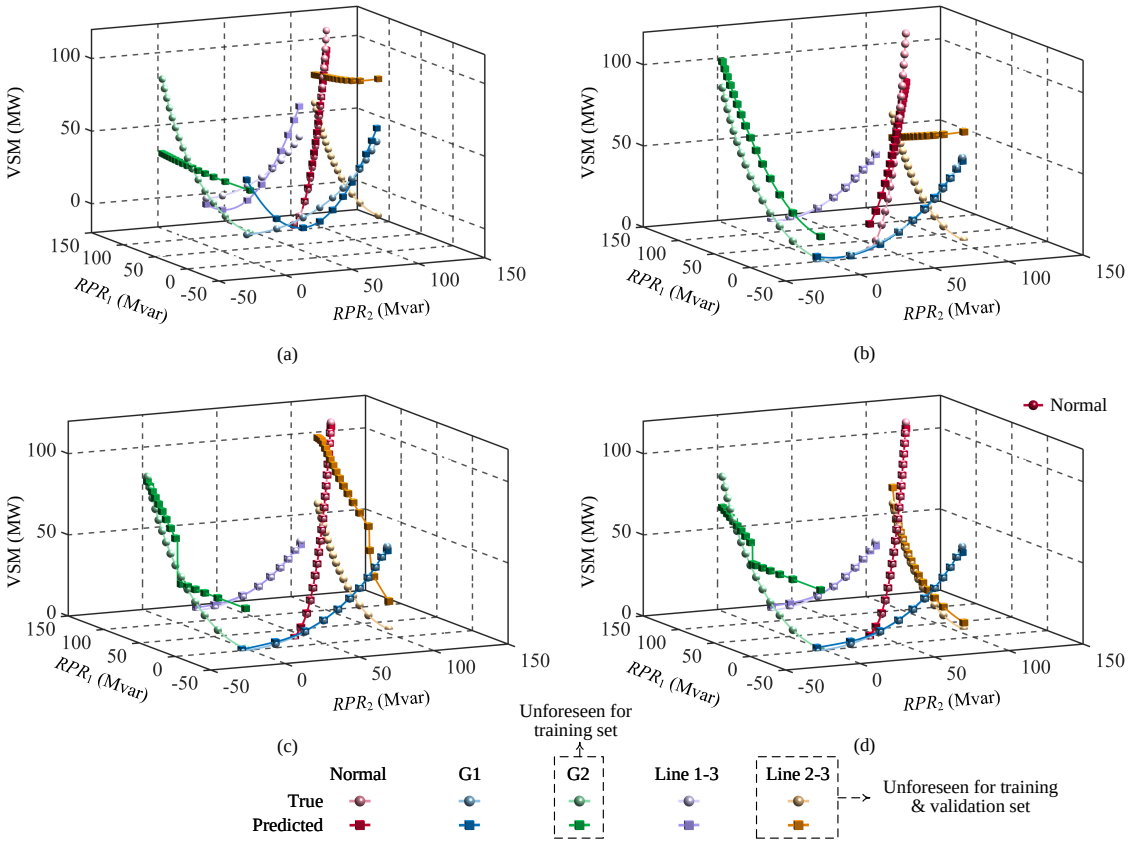


Figure 3.8: Predictions on the 3-bus system. The true VSMs, which have been shown in Figure 3.2, are plotted with light colors and sphere markers; the corresponding predicted values are plotted with dark colors and cube markers. (a) VSMs are predicted by a single quadratic MLRM (trained by LASSO); (b) VSMs are predicted by two quadratic MLRMs (trained by LASSO); (c) VSMs are predicted by proposed local LASSO regression; (d) VSMs are predicted by proposed local LASSO regression, after the database is augmented with 100 random operating points under Line 2-3 outage (generated by applying random perturbations on the load). The training set only contains data of Normal, G1, and Line 1-3; the validation set contains the data of the former three and G2.

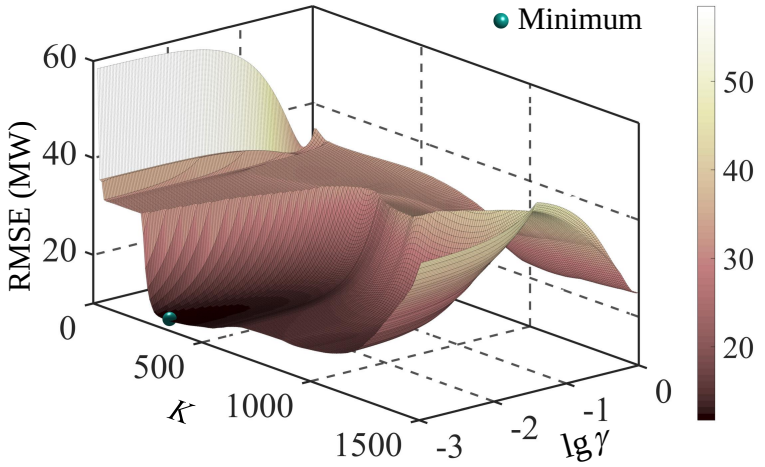


Figure 3.9: Root-mean-square prediction error over validation set when K and γ vary. The minimum (the tuned parameters) is marked by the cyan point, which corresponds to $K = 325$, $\gamma = 0.001$.

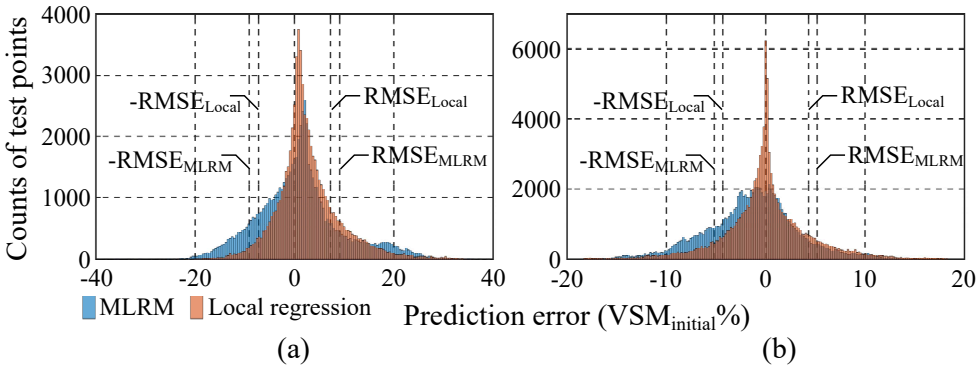


Figure 3.10: Histogram of the prediction error. (a) is for IEEE 30 system; (b) is for REAL system. Prediction error = (true VSM − predicted VSM). LASSO is used for MLRM and local regression.

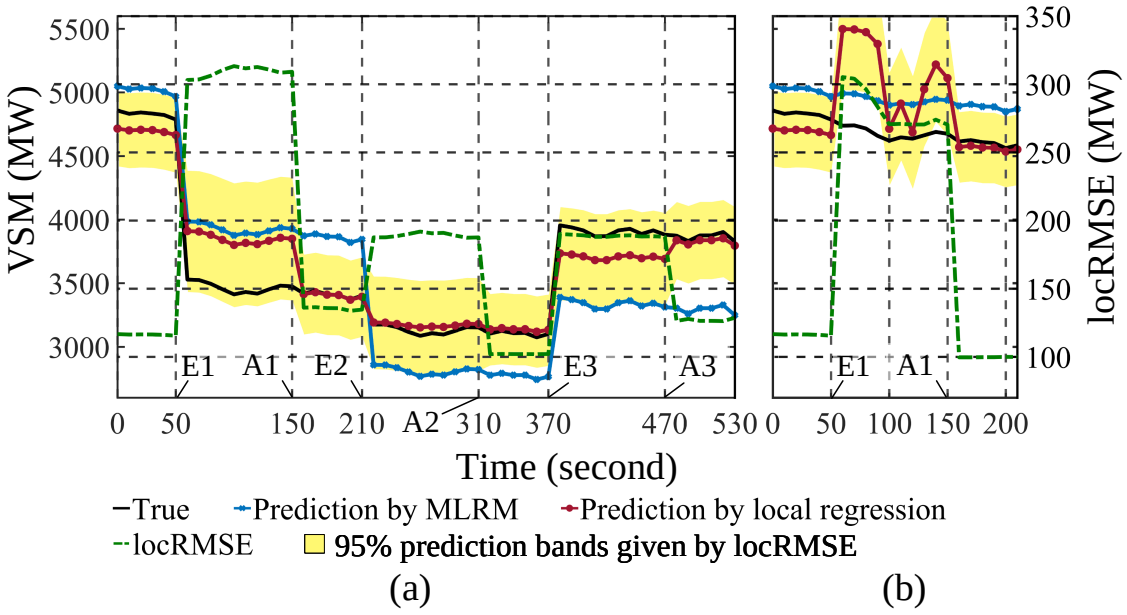


Figure 3.11: Online adapt to operating condition. (a) IEEE30 system. (b) REAL system. E: a discrete event; A: a database augmenting. The yellow area shows the prediction band given by locRMSE. To incorporate load uncertainty, the "true" VSM is the average over 40 randomly generated LIDs.

CHAPTER 4. ONLINE VSM ENHANCEMENT VIA MODEL PREDICTIVE CONTROL ON AGGREGATED THERMOSTATICALLY CONTROLLED LOADS

In order to maintain the gird stability and security, when VSM, which is monitored in real time by the proposed approach, is observed or predicted to be lower than the predefined threshold, prompt control actions are needed to steer the system back to a secure operating condition. This chapter addresses the problem of online voltage stability control under emergency.

4.1 Introduction to VSM Control by RPR

The influence of reactive power reserves in maintaining adequate voltage control and stability is widely known [5, 12, 6]. In the United States, NERC has issued standards that require transmission operators to maintain reactive resources to be used in case degenerative system conditions occur [21, 22]. Section 2.2.1.1 reviewed the voltage stability enhancement methods based on RPR management. To determine the most effective control actions that reestablish critical RPRs and VSM during emergency conditions, [113] proposed a consecutive quadratic programming framework which explicitly constrains VSM while minimizing the control cost. It is referred to as the *reference approach* in this study. In this section, we first briefly describe the reference approach and its limitations, then discuss the opportunities to overcome these limitations, which leads to our new proposed approach.

4.1.1 Formulation of The Reference Approach

The reference approach formulates the VSM control problem as follows:

$$\begin{aligned} \min_{\Delta u} \quad & \Delta u^\top \boldsymbol{R} \Delta u \\ \text{s.t.} \quad & x + \boldsymbol{S}_R \Delta u \geq x_{\min} \\ & M + \alpha e^T \boldsymbol{S}_R \Delta u \geq M_r \\ & V_{\min} \leq V + \boldsymbol{S}_V \Delta u \leq V_{\max} \\ & u_{\min} \leq u + \Delta u \leq u_{\max} \,, \end{aligned} \tag{4.1}$$

where $u = \left(P_G^\top, B_{shunt}^\top, P_{load}^\top, Q_{load}^\top\right)^\top$ is the vector of control variables containing generation dispatching, shunt switching, active and reactive power load controls; $\boldsymbol{R}$ is a positive definite matrix containing the cost weights of all control variables; x is the vector of selected PRRs which contribute the most to the voltage collapse in voltage security assessment (note: in this chapter we will use x to denote RPR column vector for convenience, whereas in chapter 3 x is the row vector of RPR observation); $\boldsymbol{S}_R = \partial x / \partial u$ and $\boldsymbol{S}_V = \partial V / \partial u$ are the sensitivity matrices of x and V with respect to control variables, which are implicitly determined by power flow equations; $e = (1, 1, \ldots, 1)^\top$; and α is a scalar parameter which reflects the approximate linear relationship between M and the total critical RPR. Problem (4.1) is a convex quadratic programming problem, which can be solved efficiently and reliably.

Controls are triggered immediately when any constraint violation is detected. After the optimum of problem (4.1) is solved and applied at time t, constraints may still not be satisfied at time $t+T$ (due to linearization error), where T is the period of control. Thus, several rounds of control actions, or *iterations of control*, may be necessary to satisfy all constraints. Numerical tests showed usually a few (3∼4) rounds of control actions are needed to meet the constrains.

In the above formulation, the relationship between M and x is simplified by

$$\Delta M = \alpha \sum_i \Delta x_i \,. \tag{4.2}$$

The value of α is set to a positive large number at the first iteration of control to prevent M constraint from binding, and then updated according to equation (4.2) in subsequent iterations of control, where ΔM and $\sum_i \Delta x_i$ are supposed to be both monitored by EMS.

4.1.2 Limitations of The Reference Approach

There are several limitations in the reference approach.

First, it uses an *unnecessarily coarse approximation* of margin sensitivity. Notice that (4.2) ignores the discrepancies among the selected RPRs – all RPRs are equally weighted in contributing to VSM. In other words, the VSM sensitivity with respect to RPRs are assumed to be $\alpha[1, 1, \ldots, 1, 0, 0, \ldots, 0]$, where the first several unit elements correspond to the selected critical RPRs. This is obviously not the case in reality. And more importantly, it does not fully utilize the RPRs-VSM relationship provided by the VSM monitoring tool, where the weight of each RPR contributed to VSM has been reflected by its coefficient in the predictive model. Thus, the optimality of control is sacrificed.

Second, the control decision is *shortsighted*. The framework does not have a look-ahead consideration (decisions are made only based on the difference between the current state and the target), which leads to several unfavorable consequences. First, it cannot predict the post-control operating condition, even though a sequence of control actions is necessary to achieve the objective. For example, loads are naturally evolving, followed by the variation of generation dispatch. However, they are not considered in the framework, even though they are predictable, and influential to VSM. This may lead to a suboptimal control, or in the worst case causing the oscillation of operating condition. Second, intertemporal constraints cannot be well established in such a framework. For example, there may exist the limits on ramping rate of generations, the limits on the frequency of switching, and the limits on the amount of load reduction with certain period. All of these limits cannot be properly considered by an optimization without look-ahead consideration.

Finally, the formulation is *self-centered* in the sense that it ignores many important external constraints and the behaviors other relevant controllers. For example, power balance and network

constraints are not explicitly considered while active power loads and generations are both supposed to be fully controlled. Consequently, the control is probably inapplicable, or the control effect can be overturned after the emergency is eliminated (generations given by economic dispatch or real-time market on normal condition could be quite different from those given by the VSM control in emergency, and therefore could draw VSM below the requirement again). Another over-optimistic assumption that can lead to inapplicable control is that the load reduction can be freely applied by VSM control. In fact, load shedding is usually regarded as one of the last resorts that saves the system from large-area blackout, so they can only be applied when specific standards or criteria are satisfied [140, 141]. In today's power system operation, these standards or criteria usually do not include VSM requirement. Last but not least, the VSM controller is probably not granted the priority to overwrite all other operating controllers, but the reference approach does not provide the flexility to consider the coordination with other controllers.

4.1.3 Objective of The Proposed Approach

To address the shortcomings of the reference approach discussed in the previous section, and based on the discussion in 1.1.6, 1.1.7, this work proposes an online optimal control strategy for ISO control center that (i) explicitly involves VSM constraint based on the local predictive model given by (3.13) instead of the second constraint in (4.1), (ii) predicts the behaviors of other relevant controllers and the evolvement of operating condition via looking ahead, (iii) engages in a more flexible demand control measure, and (iv) fits in the time framework of economic dispatch or real-time market (solve and apply the optimal control in *near-real-time*, i.e., every 5–15 minutes). The proposed approach is expected to be a novel VSM online control strategy which is effective in an uncertain, non-stationary, and interactive operating environment.

4.1.4 Sketch of The Research Tasks

Figure 4.1 gives the preview of the major parts of this chapter. It sketches the research problems and the main techniques we proposed to resolve the problems. Particularly, after we obtain the VSM

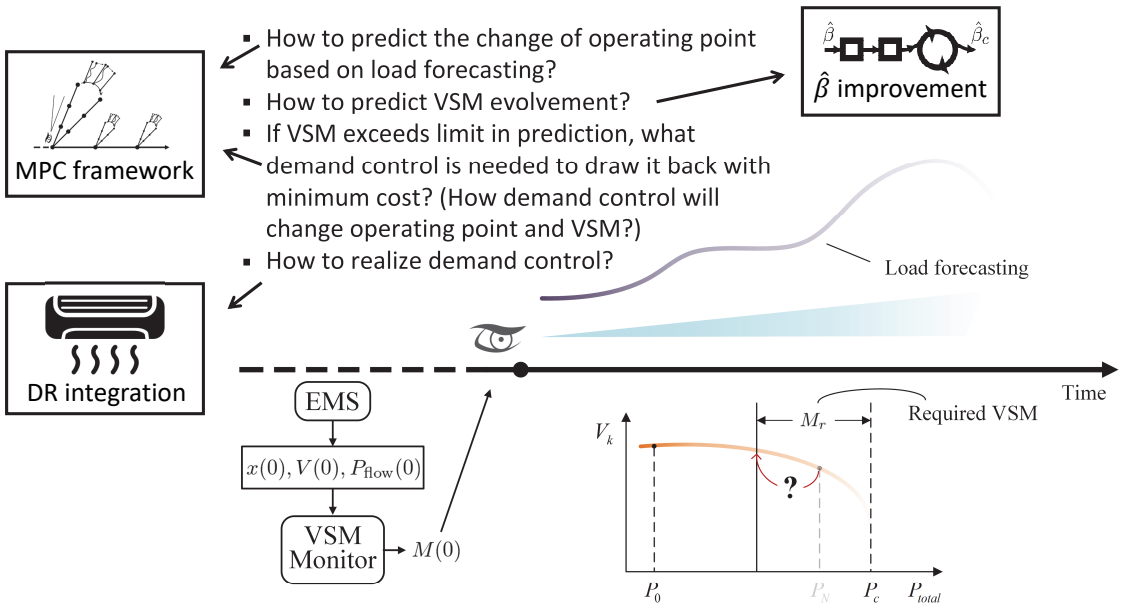


Figure 4.1: A sketch of the research tasks. The PV curve is based on the load increase pattern given by the load forecasting, where P_0 corresponds to the current operating point, and P_N corresponds to an operating point in the prediction horizon. The three blocks correspond to the following three sections.

prediction using the proposed VSM monitoring tool, we need a *proactive* controller that can look ahead and make a plan for the near future. Load evolvement can be obtained from load forecasting. But how can we predict the operating point and the VSM based on the load forecasting? If VSM exceeds the limit in this prediction, what is the optimal control action to steer it back (see the PV curve)? Finally, how do we realize the optimal control action to the real system?

The following three sections answer these questions. Section 4.2 makes necessary improvements on the VSM sensitivity given by the VSM predictive model, which enables VSM prediction for control. Section 4.3 engages in demand response (DR) through DR aggregator as an effective and economical demand control measure. Section 4.4 introduces model predictive control (MPC) as a flexible framework to design the look-ahead optimal controller considering the evolvement of the operating conditions.

In the rest of the chapter, section 4.5 summarizes the control procedure based on the proposed techniques, then the approach is demonstrated in 4.6, followed by a brief conclusion in 4.7.

4.2 Local Linear Formulation of VSM

To design a VSM state-triggered feedback controller, the first step is to build an observer, i.e., a VSM monitoring tool that can quickly evaluate VSM as well as its sensitivity information to support gradient-based optimization techniques. However, one inherent challenge for VSM-constrained optimization or optimal control problem is that VSM is intractable in representation and evaluation. Section 3.1.2 discussed the difficulty involved in expressing VSM in an explicit form.Most of the existing methods for this purpose and concluded that they do not fit the goal of designing of a near-real-time controller.

The reference approach tackled the problem by equally weighting the selected RPRs, such that the VSM sensitivity vector with respect to selected RPRs is simplified to a scalar VSM sensitivity with respect to the total selected RPR. As mentioned in 4.1.2, this simplification is unnecessary and weakens the optimality. In fact, when the online VSM monitoring tool proposed in 3 is applied, a finer approximation of VSM as a linear function of RPRs, i.e., the local VSM predictive model (3.13), is available. Let's repeat (3.13) here for convenience:

$$\hat{M}_0 = \boldsymbol{w}(z_0)\boldsymbol{M} + x_0'\hat{\beta}_0\,. \qquad \text{(3.13 revisited)}$$

$\hat{\beta}_0$ weights the contributions of RPRs on VSM, assuming the operating point is sampled from the population of the training database.

However, directly using this predictive model for control may cause problems. The next sections, 4.2.1, 4.2.2, and 4.2.3, introduces the three key modifications on the VSM predictive model that improve its performance in control.

4.2.1 Shrink The Effective Domain of Control Action via PCA

First, we observed that when there exists collinearity among the RPRs, the coefficients may lead to unreasonable control actions. This issue can be shown in Figure 4.2. Without loss of generality, suppose there are two RPRs selected as the VSM predictors, and let $\boldsymbol{X}' = [\boldsymbol{X}_1', \boldsymbol{X}_2']$ denote the design matrix for linear regression, $\boldsymbol{M}'$ denote the VSM vector as training labels. Notice that

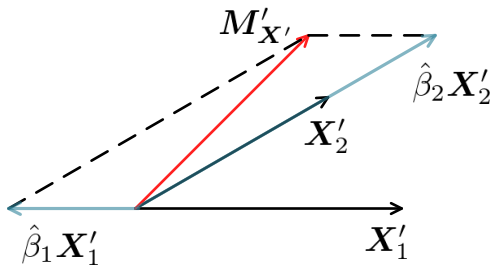


Figure 4.2: Illustration of the problem caused by collinearity.

they have been standardized to have zero means (see (3.10) and (3.11) for details). The figure is plotted on the plane spanned by $\boldsymbol{X}'_1$ and $\boldsymbol{X}'_2$. $\boldsymbol{M}'_{\boldsymbol{X}'}$ (red) denotes the orthogonal projection of $\boldsymbol{M}'$ into this spanned subspace. Linear regression using OLS simply seek the linear combination of $\boldsymbol{X}'_1$ and $\boldsymbol{X}'_2$ that equals to $\boldsymbol{M}'_{\boldsymbol{X}'}$. The coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ can be easily obtained in the figure using parallelogram law, and the blue vectors shows the two components corresponding to the basis $\boldsymbol{X}'_1$ and $\boldsymbol{X}'_2$. LASSO penalizes the L_1 norm of the coefficient vector, which could push one or both coefficients to zero depending on the regularization factor (see 3.2.2). But that is not always the case (in fact this is unfavorable for control since it implies the control actions will only target on one RPR for VSM control). Here let's consider the situation where both coefficients are nonzero. In this case, the blue vectors in Figure 4.2 will shrink with some extent, but $\hat{\beta}_1$ and $\hat{\beta}_2$ will not change the signs.

Notice that $\hat{\beta}_1$ is negative. It correctly reflects the collinearity between the two RPRs in the sample population of the training database. And it is valid when two RPRs change coherently as they usually did in the training database (usually they change in the same direction – all RPRs decrease when loads increase). However, in the control problem, the two RPRs can be manipulated independently, since the number of control variables (the controllable loads in this work) is usually much larger than the number of RPRs. Thus, when $\boldsymbol{X}'_1$ can be controlled independently, $\hat{\beta}_1 < 0$ means one can increase VSM by (only) drawing down the RPR of the first generator, which is obviously unreasonable and could cause the VSM changes in the unexpected direction.

To resolve this issue, PCA can be applied on $\boldsymbol{X}'$ before regression to eliminate the collinearity among the predictors. Then, a local regression, separated from the one for VSM prediction but through the same techniques, is conducted on the PCs exclusively to obtain the coefficients for control purpose.

However, if all PCs are preserved as predictors (no feature selection), PCA is simply applying an orthogonal transformation on the coordinate system, which indeed changes nothing physically. Based on (3.13), denote the linearized relationship between VSM and the control actions as

$$\Delta M = \hat{\beta}^{\top} \boldsymbol{S}_R \Delta u \,, \tag{4.3}$$

where $\boldsymbol{S}_R$ is the sensitivity matrix of RPRs with respect to the control variable vector u. It is trivial to show that without feature selection, (4.3) is invariant after PCA, indicating that the potential of unreasonable control actions still exits. Therefore, the problem observed in Figure 4.2 essentially is not caused by the collinearity among predictors.

Instead, the essential reason is that the control action could violate the assumption on the sparse distribution of operating points. For VSM prediction, this underlying assumption plays an important role. Notice that the number of RPRs is usually much smaller than the degree of freedom (DOF) of the operating state. Thus, strictly, VSM cannot be uniquely determined by RPRs (see 3.1.2 for details). In this case, how can we find a predictive model that maps RPRs to VSM? The rationale is based on the fact that the practical operating points of power system are statistically distributed on a manifold with a much smaller dimension than the space of operating points. In other words, the states of an operating point, including the states of control variables, are highly correlated in a statistical sense. In addition, when predicting the VSM for the current operating point, we assume it is a sample from the population represented by the training database. Therefore, the correlation among the operating states can be assumed to be also valid for the current operating point. It is this assumption that allows us to predict VSM using much less predictors (less than the true DOF of the operating state).

However, when the control is to be determined for the current condition, the control action is a *decision*, instead of a random variable that subjects to certain distribution. It is not constrained

by the sparse population distribution assumed for VSM prediction, and has its full DOF. The contradiction occurs when one is making decisions to achieve a specific consequence by assuming he will act like before. Metaphorically speaking, suppose Nick usually wears red shirt when he is happy, and wears black when he is not. In addition, he usually buys two burgers for lunch when he is happy, whereas just one when he is not so happy. Thus, we can predict today's lunch bill for Nick according to whether he wears red or black in the motoring. If Nick wants to minimize today's lunch bill, the rational decision would be directly buy one burger for lunch. The story will be ridiculous if Nick decides to minimize his lunch bill only through wearing a black shirt, or making himself unhappy (obeying his behavior pattern in history). In conclusion, *prediction can utilize correlation, but control should follow causality.*

Correlation sometimes could become (or has the same effect of) causality when the decision/behavior space is restricted. In the Nick's story, suppose he signed a contract with the burger shop saying that (i) to save time, a cashier can directly deduct money of two burgers from Nick's account when he see Nick is in red, and (ii) extra two dollars will be charged for canceling that quick order. Then, the rational decision for Nick to reduce bill, is to wear a black shirt. By signing the contract, the *effective* domain of Nick's behavior (the set of *rational* behaviors for the sake of saving money) is shrunk (e.g., he will never wear a red shirt then buy one burger, even though this action is still a feasible one).

Similarly, we can resolve the issue in the control problem to a great extent, by shrinking the effective domain of control action via PCA feature selection. Figure 4.3 shows a typical scatter plots of VSM versus all six PCs of an operating point within the neighborhood, for IEEE 30-bus system. It seems that the first two PCs (explained 87% of the predictor variance) show physically explainable linear correlation with VSM. Thus, if we shrink the effective domain of control actions to

$$\Delta\mathcal{U} = \{\Delta u \,:\, \boldsymbol{S}_R \Delta u \in \text{Range}[\boldsymbol{C}_s]\} \,, \tag{4.4}$$

where $\boldsymbol{C}_s$ is a matrix containing the first one or two columns of the PCA coefficient matrix (also called loadings) and Range$[\boldsymbol{C}_s]$ is its range space, then it will be less likely that the control action

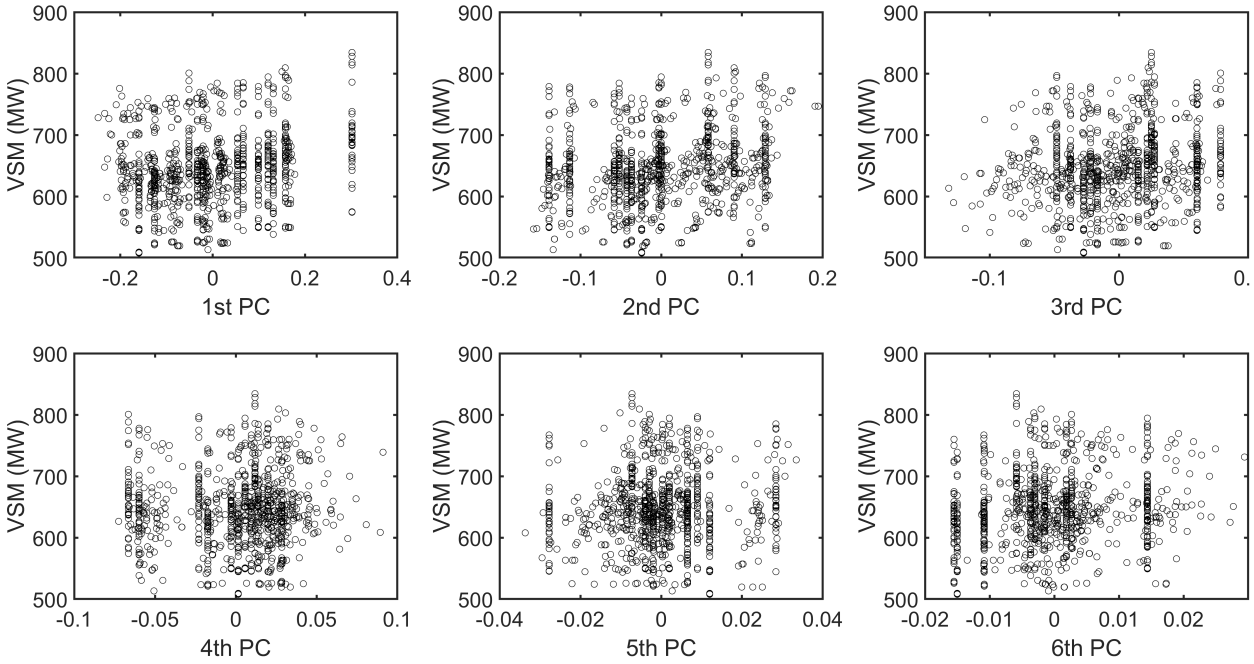


Figure 4.3: A typical scatter plots of VSM versus all principle components (PCs) for the neighbors of an operating point.

causes the VSM changes in an unexpected direction. Because in most cases, the control action that causes the simultaneous decreasing of nearly all RPRs (i.e., RPR changes that fall into Range$[\boldsymbol{C}_s]$) does reduce VSM, and vise versa. The shrinkage can be realized by only using the first one or two PCs as the predictors in the local regression. Denote the result by $\hat{\beta}_{PC}$. Then we have

$$\Delta M = \hat{\beta}_{PC}^{\top} \boldsymbol{C}_s^{\top} \boldsymbol{S}_R \Delta u = \hat{\beta}_c^{\top} \boldsymbol{S}_R \Delta u \,, \tag{4.5}$$

where

$$\hat{\beta}_c = \boldsymbol{C}_s \hat{\beta}_{PC} \tag{4.6}$$

is the coefficient vector used in control. As a result, the components of $\boldsymbol{S}_R \Delta u$ that are perpendicular to Range$[\boldsymbol{C}_s]$ are ineffective (although they are still feasible) for controlling VSM, and will be shrunk since $\|\Delta u\|$ is penalized in the objective of the optimal control. Admittedly, shrinking the effective domain of control actions sacrifices the optimality of control, but as discussed above, it is a necessary

sacrifice to obtain a reliable control, when the sensitivities of VSM with respect to control actions cannot be directly acquired in real time.

4.2.2 Equally Weight The Neighbors

The proposed VSM monitoring approach involves the kernel weights of neighbors to leverage the locality of underlying pattern, and its effectiveness has been verified [131]. However, the objective of many local regression techniques including the kernel weights, is to give an accurate prediction (zero-order information of the underlying model) at the neighborhood center, exclusively for the given input. Essentially, they are not responsible to give an accurate representation of the response sensitivities with respect to the inputs (first-order information). In other words, the purpose of those techniques is more predicting than explaining. However, the validity of the model coefficients within a neighborhood is critical for the control problem, where the operating point is moving in the neighborhood and guided by the model coefficients. This disagreement of goals sometimes leads to unfavorable control actions.

Figure 4.4 shows an example. From the scatter plot we can clearly see that VSM is increasing with the first principle component, and this is verified by the coefficients given by OLS or LASSO with equal weights. However, when kernel weights are applied, the sign of the coefficient is reversed due to the distorted empirical distribution. In this case, the weighted local model gave a more accurate VSM prediction at the center, but failed to capture the trend of VSM in the neighborhood. Consequently, the control action based on this coefficient will try to increase VSM via drawing down all RPRs, which is obviously unreasonable.

Thus, to achieve more reliable coefficients, the local regression for the control purpose should be conducted with equal weights. Besides, the regularization techniques like LASSO in our proposed approach should also be used to reduce the variance of the coefficients.

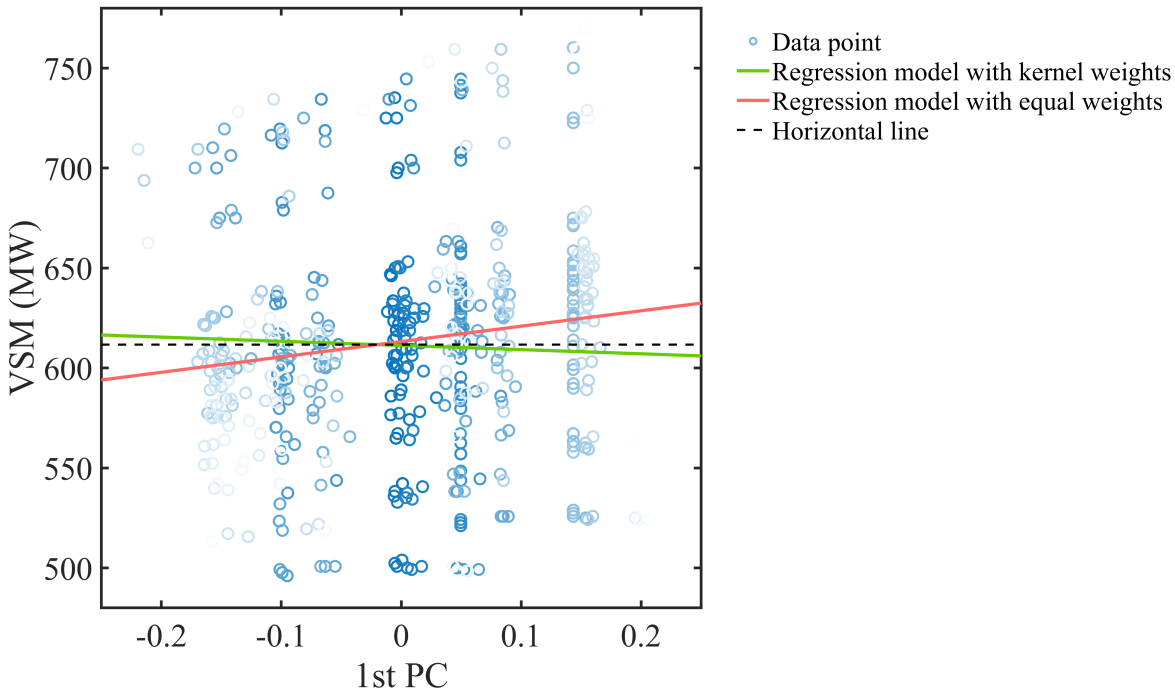


Figure 4.4: Illustration of the problem caused by kernel weights. LASSO is used in the regression. Using equal weights (red) gives an increasing model which correctly captures the data pattern. Using kernel weights (green) gives a decreasing model.

4.2.3 Correct The VSM Model Coefficients via Adaptive and Robust Re-scaling

There are many sources of the coefficient bias. Some of them, including the concept drift (the discrepancy of underlying pattern between the database and the online operating condition) and the high DOF of control, have been mitigated by our proposed approaches, but theoretically cannot be totally eliminated. To reduce their unfavorable effects, for VSM prediction problem, we focus on fully utilizing the data and the prior knowledge. But for VSM control, fortunately, we have more opportunity to do *ad-hoc* adjustment through a few of try-and-correct iterations, until the expected VSM is achieved.

The reference approach involves the response from the real power system in this try-and-correct loop – a few rounds of control action are evaluated and applied to the system until no constraints are violated. However, before the control action is applied, such a correction actually can be done using the response from the nonlinear model of the system. The nonlinear model is too complex to

be included in the online optimal control formulation, but it is available for simulating the system response. In particular, EMS has the online power flow or optimal power flow model that is able to simulate the operating point in a what-if scenario. Then, given the online measurements $z^{\text{what-if}}$ corresponding to the operating point, the VSM, $M\left(z^{\text{what-if}}\right)$, can be predicted using the proposed VSM monitoring tool. $M\left(z^{\text{what-if}}\right)$, though still a model-based prediction, is given by the nonlinear system model and VSM estimator (the predictive model (3.13) is esstially nonlinear in z_0). Thus, it is usually more accurate than the one given by the linear predictive model used in optimal control formulation (see 4.4.6), and can be used as the reference for correction.

The reference approach correct the VSM predictive model simply via re-scaling the coefficient vector. In particular, for $k = 0, 1, \ldots$

$$\alpha^{[0]} = \text{a large positive number} \tag{4.7}$$

$$\alpha^{[k+1]} = \frac{\Delta M^{[k+1]}_{\text{actual}}}{\Delta M^{[k+1]}_{\text{predicted}}} \tag{4.8}$$

$$= \frac{M^{[k+1]}_{\text{actual}} - M^{[k]}_{\text{actual}}}{\hat{\beta}^{[k]\top}\left[x^{[k+1]}_{\text{actual}} - x^{[k]}_{\text{actual}}\right]}, \tag{4.9}$$

where α is defined in (4.2); the superscript $[k]$ indicates the step of control actions; and x denotes the RPRs vector, as it is in section 3.2.1; M_{actual} can be obtained by any online VSM estimator; $\hat{\beta}$ in the reference approach, as shown in (4.2), is simply $[1, 1, \ldots, 1]^\top$; x_{actual} is online monitored.

Our proposed approach leverages the nonlinear system model and VSM estimator to do this iteration within one control step before the control action is actually applied. The VSM local linear predictive model for control with a scaling factor α can be written as:

$$M = \alpha\hat{\beta}_c^\top \Delta x + M_0 \ , \tag{4.10}$$

where $\hat{\beta}_c$ is the coefficient vector for control, given by (4.6) and with the modification described in section 4.2.2; M_0 is the VSM of current operating point. α is iteratively improved in the correction

process as follows. For a fixed control step k, and the correction step $j = 0, 1, \ldots, N_{\text{try}}$,

$$\alpha^{[k][0]} = \begin{cases} 1 & k = 0 \\ \alpha^{[k-1][\text{final}]} & k > 0 \end{cases} \tag{4.11}$$

$$\alpha^{[k][j+1]} = \frac{\Delta M^{[k][j+1]}_{\text{nonlinear}}}{\Delta M^{[k][j+1]}_{\text{linear}}} \tag{4.12}$$

$$= \frac{M^{[k][j+1]}_{\text{nonlinear}} - M^{[k]}_{\text{actual}}}{\hat{\beta}_c^{[k]\top}\left[x^{[k][j+1]}_{\text{linear}} - x^{[k]}_{\text{actual}}\right]}\,, \tag{4.13}$$

where N_{try} is a parameter that limits the number of corrections; the superscript $[k][j]$ indicates the correction step j for the control step k; the superscript $[k-1][\text{final}]$ indicates the α finally used in the $k-1$ control step; the subscript "linear" indicates the value given by the linear VSM predictive model used in control optimization; the subscript "nonlinear" indicates the value given by the nonlinear system model and the proposed VSM prediction tool. The simulated VSM on correction step $j+1$, $M^{[k][j+1]}_{\text{nonlinear}}$, is given by the following procedure: (1) use $\alpha^{[k][j]}$ as α in (4.10); (2) solve the optimal control based on (4.10); (3) on the EMS simulation platform, apply the optimal control on the current operating point (it is on control step k) and get the post-control operating point via simulation for one control time-step; (4) collect z_0 and x_0 from the operating point, and feed them into (3.13) to get $M^{[k][j+1]}_{\text{nonlinear}}$. This iterative correction process will terminate when

$$\left|\alpha^{[k][j+1]} - \alpha^{[k][j]}\right| < \varepsilon, \quad \text{or}$$
$$j = N_{\text{try}}\,, \tag{4.14}$$

where $\varepsilon > 0$ is a tolerance parameter.

On this basis, we further observed that the natural load evolvement (which usually follows the sparse distribution assumption mentioned in section 4.2.1 and can be forecasted) and the control action (which is essentially free) sometimes contribute to VSM change in different patterns. Considering this fact, we split Δx in (4.10) into two parts, and use a separate scaling factor for each to correct the coefficient vector:

$$M = \alpha_x \hat{\beta}_c^\top \Delta x_{\text{forecasted}} + \alpha_u \hat{\beta}_c^\top \Delta x_{\text{controlled}} + M_0\,, \tag{4.15}$$

where the subscripts "forecasted" and "controlled" indicate the changes due to forecasted natural evolvement of loads and the control action respectively; α_x and α_u are the two independent scaling factors. Similarly, they can be iteratively improved by

$$\alpha_x^{[k][0]} = \begin{cases} 1 & k = 0 \\ \alpha_x^{[k-1][\text{final}]} & k > 0 \end{cases} \tag{4.16}$$

$$\alpha_x^{[k][j+1]} = \frac{\Delta M^{[k][j+1]}_{\text{nonlinear, forecasted}}}{\Delta M^{[k][j+1]}_{\text{linear, forecasted}}} \tag{4.17}$$

$$= \frac{M^{[k][j+1]}_{\text{nonlinear, forecasted}} - M^{[k]}_{\text{actual}}}{\hat{\beta}_c^{[k]\top}\left[x^{[k][j+1]}_{\text{linear, forecasted}} - x^{[k]}_{\text{actual}}\right]}, \tag{4.18}$$

and

$$\alpha_u^{[k][0]} = \begin{cases} 1 & k = 0 \\ \alpha_u^{[k-1][\text{final}]} & k > 0 \end{cases} \tag{4.19}$$

$$\alpha_u^{[k][j+1]} = \frac{\Delta M^{[k][j+1]}_{\text{nonlinear, controlled}}}{\Delta M^{[k][j+1]}_{\text{linear, controlled}}} \tag{4.20}$$

$$= \frac{M^{[k][j+1]}_{\text{nonlinear, controlled}} - M^{[k]}_{\text{actual}}}{\hat{\beta}_c^{[k]\top}\left[x^{[k][j+1]}_{\text{linear, controlled}} - x^{[k]}_{\text{actual}}\right]}. \tag{4.21}$$

Again, "forecasted" means holding the control (no further action on the current step) and counting only the system changes due to forecasted load changes, whereas "controlled" means freezing the natural load changes and only counting the system changes due to the control action of the current step. Thus, the procedure of iteration is almost the same as that for single scaling factor, except that in the simulation step, hold the control u to obtain $M^{[k][j+1]}_{\text{nonlinear, forecasted}}$, whereas freeze the nature load evolvement and apply the control action to obtain $M^{[k][j+1]}_{\text{nonlinear, controlled}}$. In practice, for each correction step, we first do simulation with both natural load evolvement and control action to get $M^{[k][j+1]}_{\text{nonlinear}}$. Terminate the correction process if the criteria (4.14) have been satisfied. Otherwise, do the simulation of "forecasted" case. Then the simulation of "controlled" case can be omitted by adopting $\Delta M^{[k][j+1]}_{\text{nonlinear, controlled}} \approx \Delta M^{[k][j+1]}_{\text{nonlinear}} - \Delta M^{[k][j+1]}_{\text{nonlinear, forecasted}}$.

We also observed that α_x and α_u sometimes (although not often) rise to a very large number following the correction process mentioned above. This happens when (i) the VSM predictive

model is inaccurate, or (ii) the natural load evolvement or the control action can cause a normal change of VSM, but occasionally will not impact $\Delta x_{\text{forecasted}}$ or $\Delta x_{\text{controlled}}$ too much, or (iii) the natural load evolvement or the control action can cause a normal change of VSM, but $\Delta x_{\text{forecasted}}$ or $\Delta x_{\text{controlled}}$ is almost orthogonal to $\hat{\beta}_c$. In these cases, it makes the control non-robust to the error of model and measurements and less reliable. For example, if the nonlinear system model or the VSM monitoring tool used for simulation contains some error, then the VSM change could be significantly over-/under-estimated in the control optimization; more importantly, even if they are accurate in predicting the VSM change for the current control step, they could make large errors in the future control steps in the prediction horizon, which finally makes the control optimization infeasible or suboptimal.

To make the scaling factors more robust, just like we use LASSO to reduce the variance of $\hat{\beta}$ and improve the robustness, here we can add a proper regularization in the correction process. Take α_x for example. Equation (4.17) can be equivalently written as

$$\alpha_x^{[k][j+1]} = \underset{\alpha}{\operatorname{argmin}} \left| \Delta M^{[k][j+1]}_{\text{nonlinear, forecasted}} - \alpha \Delta M^{[k][j+1]}_{\text{linear, forecasted}} \right| . \tag{4.22}$$

Add a regularization term on the objective gives

$$\min_{\alpha} \quad \left| \Delta M^{[k][j+1]}_{\text{nonlinear, forecasted}} - \alpha \Delta M^{[k][j+1]}_{\text{linear, forecasted}} \right| + \varrho s \tag{4.23}$$

$$\text{s.t.} \quad \left\| \alpha \hat{\beta}_c^\top \boldsymbol{S}_R \right\|_1 - s \le \kappa \,, \tag{4.24}$$

$$s \ge 0 \,. \tag{4.25}$$

where $\boldsymbol{S}_R$ is the sensitivity matrix of RPRs with respect to the loads; s is the slack variable; κ is a constant parameter; ϱ is the regularization factor; so $\alpha \hat{\beta}_c^\top \boldsymbol{S}_R$ represents the sensitivity vector of VSM with respect to the loads. For α_u, simply replace "forecasted" by "controlled" in the formulation above. Solving the problem is trivial for standard optimization solvers. To derive a meaningful κ, consider how much VSM change can be actually made by a load disturbance. According to the definition of VSM, for a specific LID, denote the VSM of the base point as

$$M_0 = \|P_c - P_0\|_1 \,. \tag{4.26}$$

where P_c is the load vector of the critical point; P_0 is the load vector of the base point. Then, when a small load perturbation δP_0 is applied to the system, P_c will move to $P_c + \delta P_c$, and M_0 will become to $M_0 + \delta M$. Thus we have

$$\begin{aligned} |\delta M| &= \left| \|P_c + \delta P_c - P_0 - \delta P_0\|_1 - \|P_c - P_0\|_1 \right| \\ &\leq \|\delta P_c - \delta P_0\|_1 \\ &\leq \|\delta P_c\|_1 + \|\delta P_0\|_1 \\ &\leq (1 + r) \|\delta P_0\|_1 \ . \end{aligned} \tag{4.27}$$

where $r = \|\delta P_c\|_1 \,/\, \|\delta P_0\|_1$. On the other hand, according to (4.10), $|\delta M|$ can be predicted by

$$\begin{aligned} |\delta M| &= \left| \alpha \hat{\beta}_c^\top \boldsymbol{S}_R \delta P_0 \right| \\ &\leq \left\| \alpha \hat{\beta}_c^\top \boldsymbol{S}_R \right\|_1 \cdot \|\delta P_0\|_1 \ , \end{aligned} \tag{4.28}$$

and the equality can be achieved for certain δP_0. Thus, if $\left\| \alpha \hat{\beta}_c^\top \boldsymbol{S}_R \right\|_1 > (1 + r)$, there will be some δP_0 that can cause the predicted VSM change exceeds its actual limit. Thus, it is meaningful to let

$$\kappa = 1 + r \tag{4.29}$$

Suppose the manifold of critical points in the load space does not change significantly due to δP_0, and the manifold is sufficiently "flat" surrounding P_c, then r should be small. In the tests we use $r = 1$, such that $\kappa = 2$. ϱ can be tuned in off-line simulations based the performance of control. In the tests we use $\varrho = 0.2$.

The scaling factors given by the optimization probably sacrifices some accuracy of the predicted VSM in control optimization, but it enhances the robustness of the predictive model.

4.3 Utilization of DR Aggregator With Customer Dissatisfaction Constraint

It has been shown that load reduction is an effective measure to promptly mitigate voltage stability issue and save the system from large-area blackouts. In fact, load and generation shedding

is one of the main measures for stability control under emergency that is running in the EMSs all over the world. However, traditional stability and security control methodologies based on load shedding are generally:

(i) Reactive rather than proactive, as they usually operate only when the operating condition satisfies certain criteria determined from offline studies [140, 141];

(ii) Parsimonious, as they usually attempt to minimize the amount of load shedding based on heuristics.

This tendency can sometimes have grave consequences for the system. Many studies [142, 143] on cascading and large blackout events demonstrated that promptly shedding a certain amount (usually small) of load at the early stage of emergency would have prevented the cascading failures and might have saved the system from a blackout.

The main reason for this tendency is that load shedding is costly. [144] estimated that the national cost of power interruptions to electricity customers of the U.S. in 2015 is $59 billion, and the average cost per MWh annual sales for commercial customers is $31.

These facts motivate people to seek the alternative demand reduction strategy to reduce the cost of the traditional load shedding. Demand response (DR) provides ample opportunities to achieve a more flexible demand control and significantly reduce the cost.

4.3.1 DR and DR Aggregator

In DR programs, customers are getting incentives to allow their loads being controlled as specified in their contracts or in the DR program specific rules. The contract also involves the control restrictions and limitation such as the number of load interruptions, time limitations, and the comfort limits. Load-acceptability is facilitated by connecting the end-users loads with fast monitoring and control infrastructure. The smart grids and the recent advancements in communication and control systems at the distribution side, such as the Advanced Metering Infrastructure (AMI) and the programmable thermostats, facilitate the deployment of the DR control and make it feasible even over a wide geographical areas [145]. The most prevailing loads at the distribution networks

are the Thermostatically Controlled Loads (TCLs) such as air-conditioners, space heaters, and water heaters [146]. The large number of these devices makes the majority of energy consumption stems from the TCLs operation.

[143] developed an extended Markov model (EMM) to represent the aggregated dynamic behavior of a large number of distributed TCLs. A MPC framework is also developed based on the EMM to sculpt a *DR aggregator* that controls the aggregated TCLs at the substation level. It utilizes the ON/OFF switching capability to determine the optimal sequential set-point control law and curb down the power oscillations. Equipped with the DR aggregator, the aggregated load at the substation can track the reference power given by ISO control center with neglectable delay and oscillations, thus becomes a dispatchable source for ISO.

Using DR aggregator can significantly reduce the cost of load reduction. First, it only impacts TCLs whose effect (e.g. the in-door temperature) will not be significantly changed or even not be noticeable due to the brief period of outage. More importantly, the DR controller that combines ON/OFF and set-point controls can further reduce the loss and interruption caused by TCL reduction by means of (i) minimizing the ON/OFF times which is a major factor of the life expectancy of devices, and (ii) eliminating the abrupt change or oscillation of power caused by naive ON/OFF control (e.g. the cold load pickup issue). As a result, the stability and security control based on the DR aggregator can be more flexible and effective in the sense that (i) load reduction can be conducted promptly or even preventively (proactive control), and (ii) the amount of load reduction can be more aggressive to guarantee the emergency is reliably eliminated.

4.3.2 Introducing The Constraint of Customer Dissatisfaction

The overarching goal of the DR aggregator is to develop a framework to efficiently coordinate participation of a large number of distributed DR resources present in the hierarchical network. The large quantity and diversity of the TCL behaviors under a substation are taken care of by a DR aggregator, thus basically hidden and decoupled from the ISO central controller. Benefitting from this hierarchical control decomposition, from long-term point of view, the central controller

at ISO controls the aggregated TCLs through DR aggregator is almost like controlling a single dispatchable power source. For each control step, the aggregators of all substations first submit their applicable ranges of reference power to the central controller. The central controller then determines the expected loads at the substations and send them as the reference commands to the aggregators. Finally, the aggregators execute the commands and achieve the reference powers within one control time-step.

Unfortunately, not all inner constraints of TCLs can (or should) be hidden from the central controller or simplified as an instant (only related to the current control step) feasible range of the reference power. In this work, we particularly consider the constraint of customer dissatisfaction as an example.

The biggest advantage in utilizing the TCLs is the inherent thermal storage capability. Due to the inertia of the thermal storage, short period power outage is even unnoticeable to the customer. However, sustained reduction of the power demand will finally discomfort a customer. In the context of TCL control, the in-door temperature deviation from the customer's preference, or an increasing function of this value (e.g., square), is widely used in the literature to measure the "discomfort level" of a customer. But from the perspective of human perception, within a reasonable range, it is the cumulative discomfort temperature that really matters to the customer. For example, a reasonable description of a customer's temperature tolerance could be "within one hour, someone can tolerate 100 °F for 10 minutes, 90 °F for 15 minutes, 80 °F for 30 minutes, and 75 °F for 60 minutes". People are annoyed if their preferred temperature setting can never be achieved even though the deviation is not big, and possibly takeover the control from the aggregator in that case. Thus, the real constraint of customer discomfort due to TCL control is time-related, i.e., an *intertemporal* constraint. In order to correctly reflect this fact, we use the cumulative discomfort, the time integral of discomfort level, to measure the negative utility of DR participation, and call it *customer dissatisfaction*. This idea is implied in the literature [143].

Another fact related to the customer dissatisfaction is that it is naturally released with time. For example, people almost "forget" the discomfort temperature experienced in the previous day.

The real mechanism of forgetting is a complex science problem. In this work, we simply model it as a continuously exponential receding process that can be mathematically described as a first order ordinary differential/difference equation (ODE).

Notice that the time framework related to the customer dissatisfaction may range from minutes to hours as in the above example. But the time framework of the DR aggregator's MPC is on the seconds level. That implies the customer dissatisfaction constraint cannot be directly tackled by the DR aggregator. That is to say, DR aggregator's MPC with only a minute prediction horizon, cannot give a preventive feasible range of the reference power, taking into consideration the intertemporal constraint over an hour. Thus, this constraint must be respected and managed by the ISO central controller whose time formwork can reaches one hour or more.

In order to be considered in the central control, the customer dissatisfaction must be aggregated at the substation level. This should be a routine of DR aggregator. The DR aggregator proposed by [147] can aggregate the discomfort level and the customer dissatisfaction though the weighted (based on thermal capacity) average of individual customers. In this work, we assume the DR aggregator has such a capability to aggregate the discomfort level and the customer dissatisfaction.

In this work, we simply use temperature deviation from the preference to represent the *discomfort level*. For the ith aggregator, Denote the aggregated discomfort level as $T_{d,i}$ in °C, and denote the aggregated customer dissatisfaction as R_i. The constraint on R_i in the predictive model of central controller can be formulated by

$$R_i(k+1) = \overbrace{\alpha_{R,i} \cdot R_i(k)}^{\text{releasing}} + \overbrace{T_{d,i}(k) \cdot h}^{\text{accumulation}} \tag{4.30}$$

$$T_{d,i}(k+1) = T_{d,i}(k) + (1-\alpha_{T,i}) \left[\eta_i \cdot P_{v,i}(k+1) - T_{d,i}(k)\right] \tag{4.31}$$

$$|R_i(k)| \leq R_{i,\max}\,, \tag{4.32}$$

where $\alpha_{R,i} \in [0,1)$ is the time constant (particularly, the forgetting factor) that controls the rate of the natural releasing of R_i; h is the control time-step length, typically 5–15 minutes; $\alpha_{T,i} \in [0,1)$ is the time constant that controls the change rate of discomfort level; $\eta_i < 0$ is a constant factor that linearly maps the load change duo to the control, $P_{v,i}$, into the equilibrium $T_{d,i}$ due to the control

(explained below); $R_{i,\max} > 0$ is the limit of aggregated customer dissatisfaction; k indicates the control step in the prediction horizon of the central controller.

Equation (4.30) defines the customer dissatisfaction R (in $^\circ\mathrm{C} \cdot \mathrm{min}$) as the combination of the two effects: the accumulation of aggregated discomfort level, and the natural releasing with time. Equation (4.31) gives the relationship between the aggregated discomfort level (T_d) and the control effect on loads (P_v) as a first order ODE, where the changing rate is controlled by $\alpha_{T,i}$ and the equilibrium is control by $\eta_i \cdot P_{v,i}(k+1)$. Several important considerations about this customer dissatisfaction model are discussed as follows.

First, a more intuitive formulation for (4.30) could be

$$R_i(k+1) = \alpha_R \cdot R_i(k) + |T_{d,i}(k)| \cdot h\,, \tag{4.33}$$

where R is unsigned, and is accumulated no matter whether the temperature is higher or lower than the preference. However, the absolute value function makes the system nonlinear. What is worse, it cannot be properly linearized due to the non-smoothness at zero. An important argument here is that we can reasonably drop the absolute value function. First, the sign of R can be properly explained: positive R means the customer dissatisfaction due to the high temperature, and negative R means that due to the low temperature. (4.32) constrains it from both sides (assume the limit is symmetric). Second, except the sign, (4.30) is the same as (4.33) when T_d does not change sign (one-way accumulation). The two formulations differ the most when T_d frequently changes sign. However, in our application, this should rarely happen because: (i) the demand of load control, either for decreasing or increasing the load, is usually one-way in a period of minutes to one hour (typical prediction horizon of central controller); (ii) the square of T_d is penalized in the objective (4.52) (described in 4.4), which inherently discourages variation of T_d. One exception happens when the central MPC controller predict that R will start to increase in the near future (e.g., the peak load hour is approaching). In this case, (4.30) encourages to pre-cool the rooms, since the negative cumulation of T_d in advance can cancel out some of the positive cumulation in the near future. This is actually a favorable feature that intentionally endowed in many DR programs [148, 149]. Besides, in this case it makes even more sense than (4.33) from the perspective of human perception – right

after the room experienced a cold period, the temperature higher than the normal preference will first bring a favourable feeling via quickly releasing the negative dissatisfaction.

Regarding the relationship between P_v and T_d, (4.31) is the result of three assumptions. First, if $P_v = 0$ is held (free the load from DR control), the aggregated discomfort level T_d will converge to zero. Second, the change of T_d due to nonzero P_v can be approximated by a first order ODE with the time constant α_T. Third, a sustained nonzero P_v will shift T_d to a new equilibrium, and the equilibrium is proportional to P_v by a coefficient η. The third assumption is a widely used simplification in the literature. α_T and η can be estimated through off-line data analysis.

$\alpha_{R,i}$ and $R_{i,\max}$ can be obtained through a questionnaire survey to the customer when signing the DR program contract. For example, at first, ask the customer what is his/her preference temperature. Then ask for how long he/she can tolerate for $\pm 2°$C, $\pm 4°$C, ... in one hour (or any other practical resolution). This gives a set of points $\{(t_{\text{tol,n}}, T_{d,n})\}$ as the customer's tolerance description mentioned early in the example ($t_{\text{tol,n}}$ is in minitue). Then, we can approximately fit $T_{d,n}$ as a hyperbolic function of $t_{\text{tol,n}}$. That is to say, $t_{\text{tol,n}} \times T_{d,n} \approx c$. The constant c is an estimate of his/her $R_{\max}$. Find the value $T_{d,\text{sustained}} = c/60$, which means the customer roughly can tolrate $T_{d,\text{sustained}}$ for a very long time. That implies that the accumulation of $T_{d,\text{sustained}}$ is balanced by the natrual releasing effect. Thus, $\alpha_{R,i}$ can be obtained through

$$c = \alpha_{R,i} \cdot c + T_{d,\text{sustained}} \cdot h \,. \tag{4.34}$$

Finally, aggregate the parameters of customers (based on thermal capacity) to obtain $\alpha_{R,i}$ and $R_{i,\max}$ in (4.30) and (4.32).

As mentioned above, from long-term point of view, we assume the aggregator can perfectly follow the reference power assigned by ISO within one control step (once all constraints are satisfied). That is to say,

$$P_{v,i}(k+1) = P_{u,i}(k) \,, \tag{4.35}$$

where $P_{u,i}$ is the control input, i.e., the reference power given by the central controller. Thus, (4.31) can be directly written as

$$T_{d,i}(k+1) = T_{d,i}(k) + (1-\alpha_{T,i})\left[\eta_i \cdot P_{u,i}(k) - T_{d,i}(k)\right] . \tag{4.36}$$

4.4 The MPC Framework

A general description about MPC can be found in Appendix C. The looking ahead and receding horizon strategy of MPC meet the target of our control design. In particular, (i) by looking ahead, it can preventively mitigate future degenerative condition and anticipate the responses (static or dynamic) of the control environment; (ii) by receding horizon, it adapts to the changing environment and self-corrects the error imposed by the model simplification and uncertainty. Besides, linear (open-loop) MPC, finally transformed into a convex quadratic programming or linear programming, can be efficiently solved by standard solvers in real time (several seconds), even for a large system. This section describes the MPC framework that realizes our proposed control strategy.

4.4.1 Modeling The Load Evolvement

Loads are the major parameters of an operating point considered in this work. In order to represent the loads in the control formulation, we need to (i) decompose the load into fixed part and controllable part (TCLs), and (ii) describe their evolvement in the prediction horizon.

Notice that the evolvement of loads is essentially a parameter, rather than a variable in the predictive model of control. In fact, it is predicted by the external load forecasting, rather than the predictive model. However, directly using the forecasted load series as parameters will make the predictive model time-varying and complicate the formulation. In this work, we expand the state space to include the loads as state variables. In particular, the forecasted load series is fitted by a first order ODE. Then include this ODE in the predictive model of control.

Another simplification made in this work in that the power factor is fixed. In several studies [150, 151, 152], controlling the reactive power at the substation has been enabled as a function of the aggregator, through the optimal operation of smart inverter based power sources. Thus, the

proposed approach can be extended to involve the controllable power factor and get rid of this simplification.

To summarize, the loads in the predictive model can be formulated by

$$\Delta P_{f,i}(k+1) = \alpha_{f,i} \cdot \Delta P_{f,i}(k) \tag{4.37}$$

$$\Delta P_{o,i}(k+1) = \alpha_{o,i} \cdot \Delta P_{o,i}(k) \tag{4.38}$$

$$\begin{aligned} \Delta P_{c,i}(k+1) &= \Delta P_{o,i}(k+1) + P_{v,i}(k+1) \\ &= \alpha_{o,i} \cdot \Delta P_{o,i}(k) + P_{v,i}(k+1) \\ &= \alpha_{o,i} \cdot \Delta P_{o,i}(k) + P_{u,i}(k)\,, \end{aligned} \tag{4.39}$$

where $P_{f,i}$ is the fixed load (note: $P_{f,i}$ is evolving; "fixed load" is a widely used term in literature, in the sense that it is unchangeable for control); $P_{o,i}$ is the forecasted TCLs supposing no control is applied; $P_{c,i}$ is the actual TCLs containing $P_{o,i}$ and the change casued by contrl, $P_{v,i}$, so, $P_{f,i} + P_{c,i}$ is the actual total load for the ith aggregator; $\alpha_{f,i} \in [0, 1)$ and $\alpha_{o,i} \in [0, 1)$ are corresponding time constants; and

$$\Delta P_{f,i}(k) = P_{f,i}(k) - P_{f,i}(\infty) \tag{4.40}$$

$$\Delta P_{o,i}(k) = P_{o,i}(k) - P_{o,i}(\infty) \tag{4.41}$$

$$\Delta P_{c,i}(k) = P_{c,i}(k) - P_{o,i}(\infty)\,, \tag{4.42}$$

where $P_{f,i}(\infty)$ and $P_{o,i}(\infty)$, the extrapolated equilibriums of forecasted load series, are obtained during the model fitting mentioned above.

4.4.2 Cooperation With Other Controllers

As mentioned early in section 4.1.2, our controller operates in a cooperative environment where the system response to our control input depends not only on the DR aggregator, but also other relevant controllers. Thus, to endow our controller with a plug-and-play feature, the predictive model should try it best to predict the responses of these relevant controllers. In this work, as an example, we use the sensitivity matrix to predict the response of the economic dispatch. In

particular, for the load change ΔP_L, we can predict the corresponding change of generation by

$$\Delta P_G = \boldsymbol{S}_{GL} \Delta P_L \,, \tag{4.43}$$

where $\boldsymbol{S}_{GL}$ is the sensitivity matrix of generations with respect to loads. $\boldsymbol{S}_{GL}$ can be derived from the optimal solution (primal and dual) of the economic dispatch problem. We suppose $\boldsymbol{S}_{GL}$ can be acquired from the economic dispatch or real-time market module. Then, the sensitivity matrices of voltages and RPRs with respect to the loads can be derived:

$$\partial x / \partial P_L = \boldsymbol{S}_{RL} + \boldsymbol{S}_{RG} \boldsymbol{S}_{GL} = \boldsymbol{S}_R \Delta P_L \tag{4.44}$$

$$\partial V / \partial P_L = \boldsymbol{S}_{VL} + \boldsymbol{S}_{VG} \boldsymbol{S}_{GL} = \boldsymbol{S}_V \Delta P_L \,, \tag{4.45}$$

where $\boldsymbol{S}_{RG}$ and $\boldsymbol{S}_{VG}$ are the sensitivity matrix of RPRs and voltages with respect to generations respectively, (see [134] for the derivation). To further improve the prediction, the designed controller can also request a set of feasibility cuts from the economic dispatch module. These cuts can prevent the control action from violating the security or market constraints considered in economic dispatch. In our simulation tests shown in 4.6, the OPF function of MatPower [153] is used to represent the economic dispatch module of EMS.

The strategy shown here can be extended to other relevant controllers inside and outside ISO control center. All these relevant controllers, including the DR aggregator, can provide the interface parameters (e.g., $R_{i,\max}$, feasibility cuts) that really reflect their capabilities, or they can just provide reserved parameters for their private concerns (e.g., preserving for other services, hiding confidential information, etc.). The philosophy of the strategy is: instead of trying to model other controllers' behaviors as the objective reality, predict their behaviors by treating them as subjective cooperators and talk to them; instead of regarding the parameters as measurements, regard them as a contract.

4.4.3 Stability and Security Constraints

Based on (4.15), (4.35), and (4.37)–(4.42), the VSM constraint can be written as

$$M(k+1) = \alpha_x \hat{\beta}_c^\top \underbrace{\boldsymbol{S}_R \left[\tilde{\alpha}_f \Delta P_f(k) + \tilde{\alpha}_o \Delta P_o(k) - \Delta P_f(0) - \Delta P_o(0)\right]}_{\Delta x_{\text{forecasted}}(k+1)} + \alpha_u \hat{\beta}_c^\top \underbrace{\boldsymbol{S}_R \left[P_u(k) - P_c(0) + P_o(0)\right]}_{x_{\text{controlled}}(k+1)} + M(0) \tag{4.46}$$

$$M(k+1) \geq M_r \,, \tag{4.47}$$

where $\boldsymbol{S}_R$ is the sensitivity matrix of RPRs with respect to loads; M_r is the predefined VSM requirement; P_f, P_o, P_c, and P_v in (4.46) are vectors for all loads (suppose every load bus has a DR aggregator); $\tilde{\alpha}_o = \text{diag}[\tilde{\alpha}_{f,1}, \tilde{\alpha}_{f,2}, \ldots]$, $\tilde{\alpha}_o = \text{diag}[\alpha_{o,1}, \alpha_{o,2}, \ldots]$ are the diagonal matrices of time constants. The voltage magnitude constraint for load buses can be written as

$$V(k+1) = \boldsymbol{S}_V \underbrace{\left[\tilde{\alpha}_f \Delta P_f(k) + \tilde{\alpha}_o \Delta P_o(k) + P_u(k) - \Delta P_f(0) - \Delta P_c(0)\right]}_{\Delta P_L} + V(0) \tag{4.48}$$

$$V_{\min} \leq V(k+1) \leq V_{\max} \,, \tag{4.49}$$

where $\boldsymbol{S}_V$ is the sensitivity matrix of voltage magnitudes with respect to the loads; $V_{\min}$, $V_{\max}$ are the secure limits of voltage magnitudes for all load buses, typical $V_{\min} = 0.95 \times \mathbf{1}$, and $V_{\max} = 1.05 \times \mathbf{1}$, where $\mathbf{1} = [1, 1, \ldots, 1]^\top$.

4.4.4 Constraint of Control Input

As mentioned in 4.4.2, aggregators can submit a feasible range for the reference power based on several concerns. The designed controller should respect this limit:

$$P_{u,\min} \leq P_u \leq P_{u,\max} \tag{4.50}$$

where all three terms are vectors for all loads (aggregators). Meanwhile, $P_{u,i}$ cannot cause a negative $P_{c,i}$. So, based on (4.39) and (4.42) we have

$$\alpha_{o,i} \cdot \Delta P_{o,i}(k) + P_{u,i}(k) + P_{o,i}(\infty) \geq 0 \tag{4.51}$$

4.4.5 Objective Function

A normal quadratic cost function is used as the objective:

$$J = \sum_{k=0}^{H} \left[T_d(k)^\top \boldsymbol{W}_T T_d(k) + P_u(k)^\top \boldsymbol{W}_u P_u(k) \right], \tag{4.52}$$

where H is the number of steps in the prediction horizon; T_d is the vector of discomfort levels for all aggregators; P_u is the vector of control input (load reduction reference); $\boldsymbol{W}_T$ and $\boldsymbol{W}_u$ are penalty weight matrix for discomfort level and control actions respectively, and they are typically diagonal and positive definite. In the test example, the weights for T_d is proportional to the base load, and the control actions are equally weighted. $\boldsymbol{W}_T$ is normalized such that $\text{trace}(\boldsymbol{W}_T) = 1$. The relative magnitude of $\boldsymbol{W}_u$ with respect to $\boldsymbol{W}_T$ is tuned based on the control performance in a number of simulations.

The objective function reflects the negative utility (cost) for the customers participating in the DR program. The first term measures the negative utility by the in-door temperature deviation, whereas the second term measures that by the power change (mainly power reduction). So, minimizing the first term will enhance customers' comfort during the demand response, while minimising the second term leads to a less load change for DR. Based on the assumption of (4.36), T_d is the integral of P_u. So the two terms are basically consistent, but emphasizes different aspects of the customer's loss due to DR. For a case showing the divergence of the two objectives, see the discussion on the control behavior at the re-closing moment in 4.6.3.3. The negative utility is accumulated with time in the prediction horizon.

To leverage the power of looking ahead, H should not be very small (e.g., ≥ 3). Also, it should not be too large to exceed the valid range of the predictive model (considering the load forecasting confidence and the linearization error of the predictive model). In the test example $H = 6$ or 12 (30–60 minutes if $h = 5$).

4.4.6 Overall MPC Formulation

In summary, utilizing the DR aggregator, the proposed controller aims at maintaining VSM while minimizing the customers' discomfort and the load reduction (or increase) for DR in near-

real-time. the linear MPC formulation for the proposed controller is given by

$$\min_{P_u} \quad \sum_{k=0}^{H} \left[T_d(k)^\top \boldsymbol{W}_T T_d(k) + P_u(k)^\top \boldsymbol{W}_u P_u(k) \right] \tag{4.52}$$

s.t. $\forall i, k$:

$$R_i(k+1) = \alpha_{R,i} \cdot R_i(k) + T_{d,i}(k) \cdot h \tag{4.30}$$

$$\Delta P_{f,i}(k+1) = \alpha_{f,i} \cdot \Delta P_{f,i}(k) \tag{4.37}$$

$$\Delta P_{c,i}(k+1) = \alpha_{o,i} \cdot \Delta P_{o,i}(k) + P_{u,i}(k) \tag{4.39}$$

$$T_{d,i}(k+1) = T_{d,i}(k) + (1 - \alpha_{T,i}) \left[\eta_i \cdot P_{u,i}(k) - T_{d,i}(k) \right] \tag{4.36}$$

$$|R_i(k)| \leq R_{i,\max} \tag{4.32}$$

$$\alpha_{o,i} \cdot \Delta P_{o,i}(k) + P_{u,i}(k) + P_{o,i}(\infty) \geq 0 \tag{4.51}$$

$$\begin{aligned} M(k+1) = {} & \alpha_x \hat{\beta}_c^\top \boldsymbol{S}_R \left[\tilde{\alpha}_f \Delta P_f(k) + \tilde{\alpha}_o \Delta P_o(k) - \Delta P_f(0) - \Delta P_o(0) \right] \\ & + \alpha_u \hat{\beta}_c^\top \boldsymbol{S}_R \left[P_u(k) - P_c(0) + P_o(0) \right] + M(0) \end{aligned} \tag{4.46}$$

$$V(k+1) = \boldsymbol{S}_V \left[\tilde{\alpha}_f \Delta P_f(k) + \tilde{\alpha}_o \Delta P_o(k) + P_u(k) - \Delta P_f(0) - \Delta P_c(0) \right] + V(0) \tag{4.48}$$

$$M(k+1) \geq M_r \tag{4.47}$$

$$V_{\min} \leq V(k+1) \leq V_{\max} \tag{4.49}$$

$$P_{u,\min} \leq P_u \leq P_{u,\max} \,. \tag{4.50}$$

To differentiate parameters from optimization variables, all parameters are shown in blue. Equation (4.30), (4.37), (4.38), (4.39), and (4.36) together form the ODEs of MPC. This convex quadratic programming can be efficiently solved by the standard solvers, such as CPLEX and Gurobi, within one minute even for a large-scale system.

In practice, in order to ensure the feasibility of the problem, slack variables can be added to the inequality constraints while their norms are penalized in the objective. In this way, a nonzero slack variable in the solution indicates the infeasibility of the original problem, and identifies which constraint is violated. In this case, other control measures are needed to be engaged in the VSM maintenance, or the operators can adjust the limits to relax the problem.

4.5 Summary of The Control Procedure

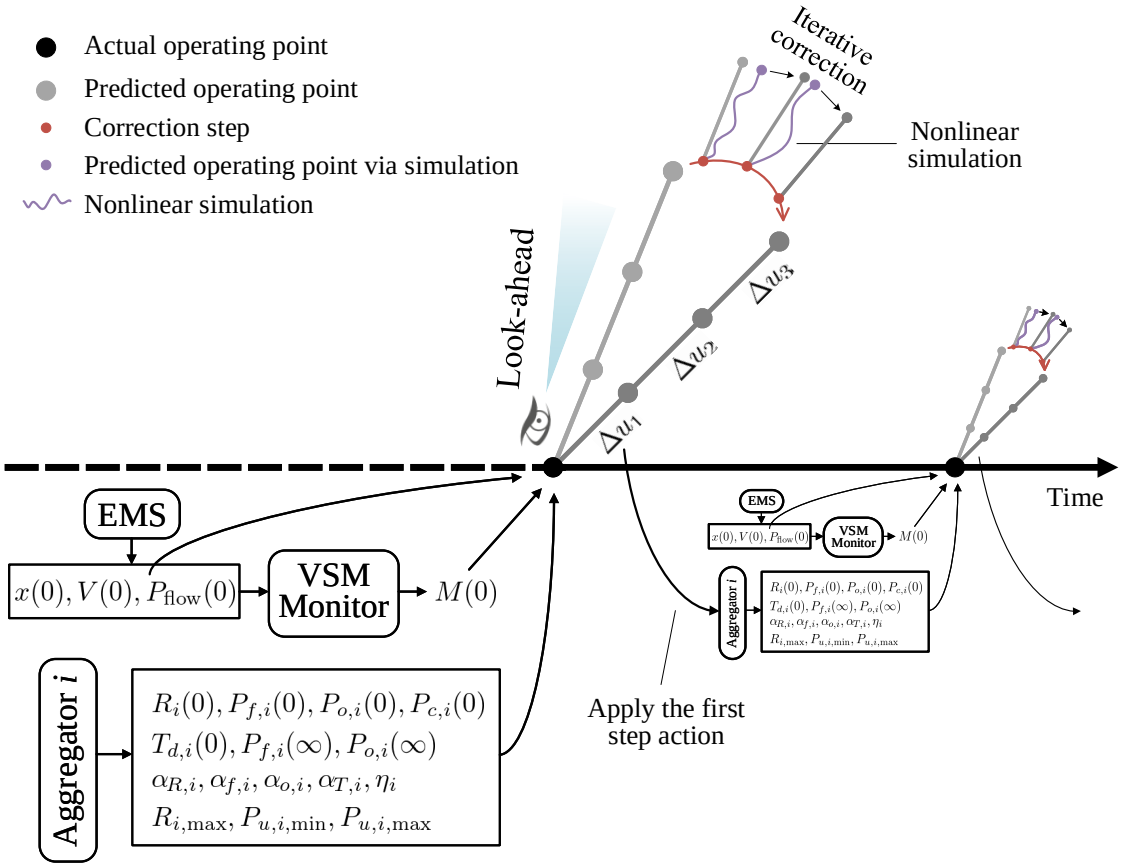


Figure 4.5: Overall control procedure. Δu_j is the control action for control step j in the MPC prediction.

The overall procedure of the proposed control approach is shown in Figure 4.6. The control is running in a receding/rolling horizon manner: (1) update information (basically the states and the knowledge of the environment); (2) make a plan for the future according to the known information; (3) implement the first step of the plan; when it is finished, again, (1) update information; ...

Particularly for the proposed control:

1. **Knowledge updating in aggregators.** At the beginning of each control step, each DR aggregator (i) collects information about its customers' in-door temperatures, temperature preferences, and dissatisfactions, then aggregates them into $T_{d,i}(0)$ and $R_i(0)$; (ii) measures

the current fixed load ($P_{f,i}(0)$), TCL ($P_{c,i}(0)$), and estimates the TCL if no DR ($P_{o,i}(0)$); (iii) based on the load forecasting, estimates $P_{f,i}(\infty)$, $P_{o,i}(\infty)$, $\alpha_{f,i}$, and $\alpha_{o,i}$; (iv) adjusts $\alpha_{R,i}$, $\alpha_{T,i}$, $\eta_i, R_{i,\max}$, $P_{u,i,\min}$, and $P_{u,i,\max}$ if needed; (v) submits all these updated values to the ISO central controller.

2. **Knowledge updating in ISO.** Meanwhile, the VSM monitoring tool predicts VSM according to the current RPRs ($x(0)$), voltage magnitudes ($V(0)$), and active power flows ($P_{\text{flow}}(0)$) given by the state estimation of EMS. The VSM prediction ($M(0)$) and the voltage magnitudes ($V(0)$) are sent to the central controller.

3. **Seeking the optimal control.** Then, the central controller (i) updates the initial conditions and the parameters obtained from the previous two steps; (ii) interactively solves the MPC problem formulated in 4.4.6 and corrects the scaling factors α_x and α_u until the stopping criterion (4.14) is satisfied; (iii) sends the first control actions in the optimal solution to the corresponding aggregators.

4. **Control execution.** Each DR Aggregator takes the control input from the central controller as the reference, runs its DR controller to implement the reference load change. The reference is held until a new value is received. When time goes to the next control period, go to Step 1.

If operating the control itself imposes considerable cost (mainly in communication and execution stage), when the system is unstressed, the proposed controller can operate in a *standby mode* as an early warning tool. In this situation, DR is not active, so the central controller does not need to know the states of customer discomfort level and dissatisfaction. Besides, the controller can use the load forecasting of control center instead of that from the aggregators. In the MPC formulation, treat all loads from the central load forecasting as TCLs (to give a nonempty feasible region for P_u). Other parameters that should come from aggregators can be set to default values. Thus, the proposed MPC will be solved without the communication to aggregators. The norm of the solution optimal control action, suspended in the standby mode, can be used as a control triggering

signal, because it should be zero when there is no constraint violation in the prediction horizon. Benefiting from the look-ahead capability, it triggers the control (switches to working mode) before the violation actually happens.

4.6 Example

The effectiveness of the proposed control approach is demonstrated on the IEEE 30-bus system. There are 6 generators (RPRs), 21 loads(aggregators), and 24 load buses (PQ buses).

4.6.1 Scenarios

Two simulation scenarios are demonstrated, emphasizing different aspects of the controller's behavior.

4.6.1.1 Scenario 1: peak load hours

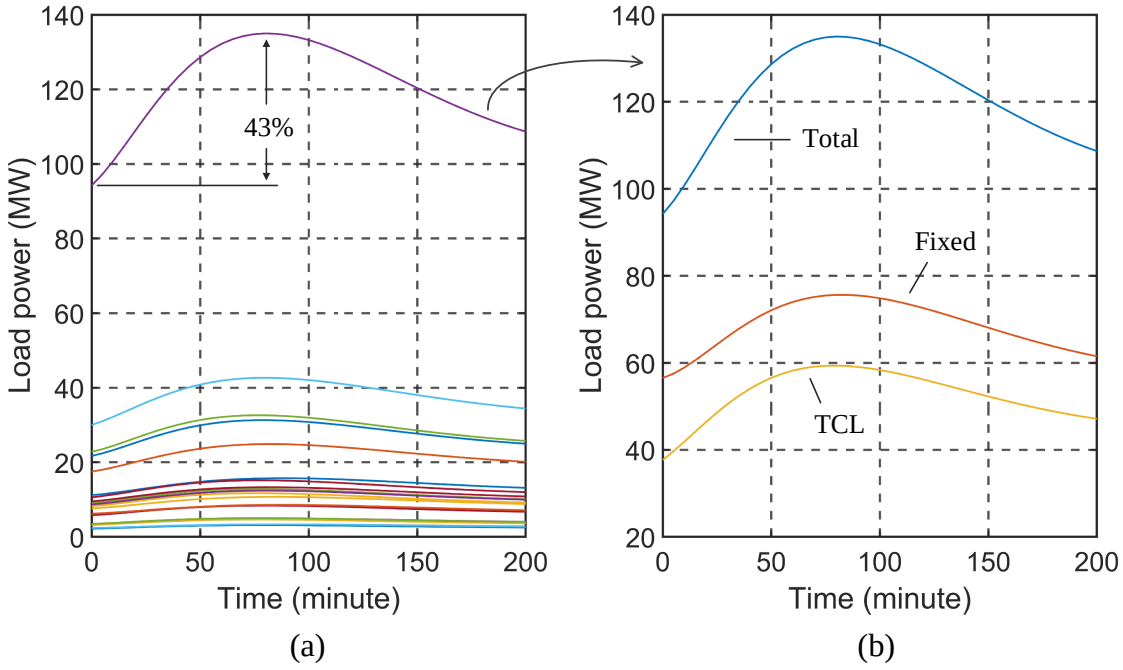


Figure 4.6: Load profiles of Scenario 1. (a) Natural evolvement of all loads (no control applied). Each curve is for one load. The total load is 283.4 MW at the initial point, and 404.4 MW at the peak. All loads change proportionally. (b) Components of the largest load ($P_{f,4} + P_{o,4}$).

In Scenario 1, the system is supposed to be operating on normal condition of a peak load day. Then the system rides through the peak load hours (200 minutes) of a roughly 40% demand increment, with the proposed DR control. The load profiles for this scenario is shown in Figure 4.6. The power factors are supposed to be fixed. Assume 40% loads are controllable TCLs (see Figure 4.6(b)). All loads change proportionally (with a small random noise). For each load, assume the changing rate of the TCL part is three times of the changing rate of the fixed part plus a small random noise (see Figure 4.6(b); the peak of the yellow curve almost doubled the base level).

The load profiles are generated through the following procedure: first set the $\alpha_{f,i}$ and $\alpha_{o,i}$ for all aggregators (0.9 plus a small random noise for different aggregators) and keep them invariant; then set the series of $P_{f,i}(\infty)$, and $P_{o,i}(\infty)$; and finally generate the load profiles shown in Figure 4.6 according to $\alpha_{f,i}$, $\alpha_{o,i}$, $P_{f,i}(\infty)$, and $P_{o,i}(\infty)$ through (4.37) and (4.38). Notice that in practice, as mentioned in 4.4.1, $\alpha_{f,i}$, $\alpha_{o,i}$, $P_{f,i}(\infty)$, and $P_{o,i}(\infty)$ are obtained based on load forecasting, and updated in every control step (provided by aggregators). But in the test here, to simplify the scenario generation process, we reversed the process (load forecasting to parameters $\rightarrow$ parameters to load profile).

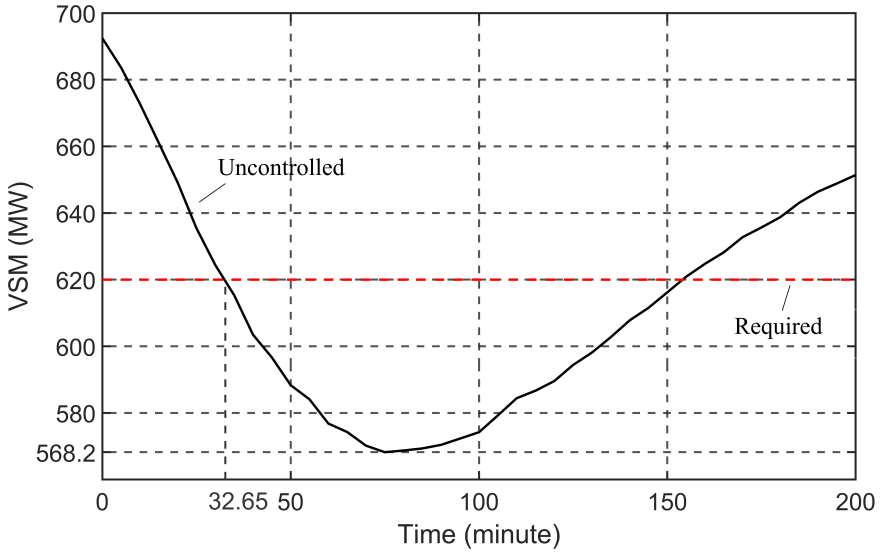


Figure 4.7: VSM profile of Scenario 1. Each point on the curve is obtained via off-line PV curve tracing in 40 randomly selected LIDs.

Given the load profiles, as mentioned in 4.4.2, the generation dispatch is determined by solving economic dispatch. Then the trajectory of operating point can be obtained. For each operating point, 40 LIDs are randomly selected (from Gaussian distribution with the base load as the mean) to conduct PV curve tracing. The average of the 40 VSMs is regarded as the true expectation and depicted in Figure 4.7. The VSM requirement is set to $M_r = 620$ MW. From the figure we can see, if no control applied, VSM will decrease to around 570 MW, and violate the VSM requirement after $t = 32.65$.

4.6.1.2 Scenario 2: peak load hours + contingency

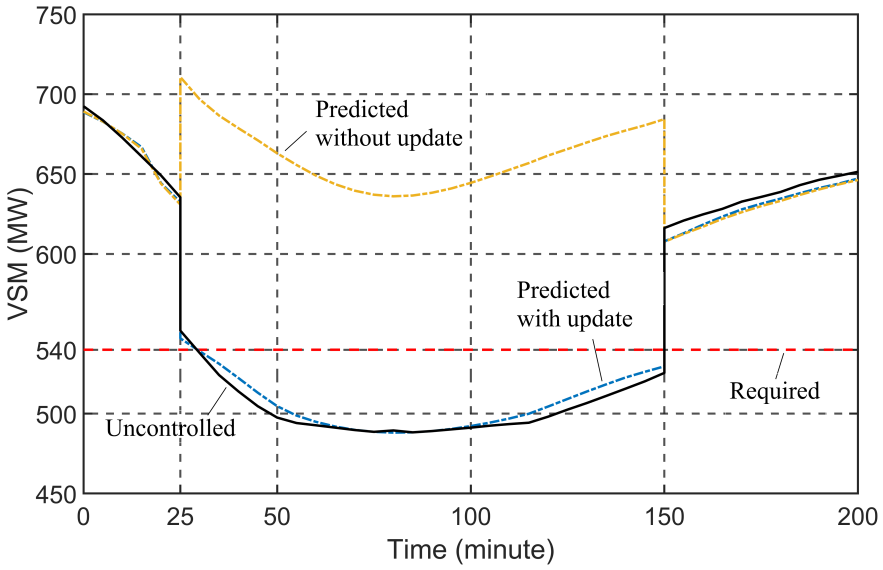


Figure 4.8: VSM profile of Scenario 2. Each point on the uncontrolled VSM (black) is obtained via off-line PV curve tracing in 40 randomly selected LIDs. The yellow dashed curve is the predicted VSM without database updating, which failed to capture the VSM under contingency. The blue curve is the predicted VSM with one round of database augmentation conducted after the contingency happens. The time needed for database updating is not shown. As a result, the prediction (blue curve) dived vertically at $t = 25$.

In Scenario 2, in order to observe the controller's behavior under sudden change of VSM, we further applied a transmission line outage during the peak time in Scenario 1. The line between Bus 6 and Bus 8 is tripped at $t = 25$, then reclosed at $t = 150$. The load profiles are identical to

Scenario I. The VSM profile without control is shown in Figure 4.8. In practice, operators will set a lower requirement of VSM under contingency. So, in Scenario 2, M_r is set to be 540 MW. The contingency dropped VSM roughly 100 MW down, thus made it close to the limit.

An important observation in this scenario is, in order to accurately track the post-contingency VSM, the database updating is needed right after the contingency happens (see section 3.3.2 and 3.5.4). Thus, the control inputs for the DR aggregators should be locked until the database updating finished (system stability in this period relies on the short-term control tools that are beyond the scope of this work). Usually, the updating can finish within one control step (100 seconds is assumed in section 3.5.4).

4.6.2 Parameters

As shown in 4.3.2, $\alpha_{R,i}$ and $R_{i,\max}$ can be obtained through a survey to the costumers, whereas $\alpha_{T,i}$ and η_i can be obtained through off-line data analysis and updated online by the aggregators. In this test example, for $\alpha_{R,i}$ and $R_{i,\max}$, we reserved that process: first set the $\alpha_{R,i}$ for all aggregators (0.875 plus a small random noise, which means 80% of the dissatisfaction will be released within an hour if T_d is fixed at zero), then use (4.34) reversely to get $R_{i,\max}$, assuming $T_{d,\text{sustained}} = 1.5°\text{C}$ plus a random noise. $\alpha_{T,i}$ is randomly selected from $(0.6, 0.7)$ and kept invariant during the simulation. η_i for the smallest load is set to 5, which means 1 MW load reduction of the smallest load (2.52 MW) will cause on average 5°C in-door temperature increment for the customers of the aggregator. η_i for other loads are inversely proportional to their base loads plus a small random noise.

$\alpha_{f,i}$ and $\alpha_{o,i}$ have been obtained during the scenario generation (see 4.6.1.1).

To fully utilize the TCLs, we set $P_{u,\min} = -\infty$. Notice that $P_{u,\min}$ is still bounded below by (4.51), the equality holds when the TCL is exhausted for DR control. For aggregator i, we set $P_{u,i,\max} = 0.5P_{o,i,\text{initial}}$ and keep it invariant, where $P_{o,i,\text{initial}}$ is the TCL at the initial operating point.

The prediction horizon H is set to 12.

4.6.3 Simulation Results

We simulated the evolvement of operating point when the proposed control is applied to the two scenarios described above. For each control step:

1. With given initial states, solve the MPC problem presented in 4.4.6, with the adaptive correction described in 4.2.3. The optimal control action for the current step is obtained.

2. Use the ODEs of MPC ((4.30), (4.37), (4.38), (4.39), and (4.36)) to simulate the one-step changes of the loads, R, and T_d. Regard the results as the true values of the initial states for the next control step (note: this is an assumption for simulation; in practice the initial states are the control feedback from the system, monitored and submitted by the aggregators as described in 4.5).

3. Solve the economic dispatch OPF problem to simulate the generation and voltage responses to the load changes. Again, regard this simulated result as the true system change. At this point, the operating point after the current control step, i.e., the initial states of the next control step, has been obtained.

4. Using RPRs, voltages, and power flows of this operating point as the inputs, predict the corresponding VSM using the proposed monitoring tool. Regard the result as the true value ($M(0)$ for the next control step).

For Scenario 1, we further show the control performance when random disturbances are added to the ODEs of MPC. That is to say, (4.30), (4.37), (4.38), (4.39), and (4.36) become

$$R_i(k+1) = \alpha_{R,i} \cdot R_i(k) + T_{d,i}(k) \cdot h + w_{i,1} \tag{4.53}$$

$$\Delta P_{f,i}(k+1) = \alpha_{f,i} \cdot \Delta P_{f,i}(k) + w_{i,2} \tag{4.54}$$

$$\Delta P_{o,i}(k+1) = \alpha_{o,i} \cdot \Delta P_{o,i}(k) + w_{i,3} \tag{4.55}$$

$$\Delta P_{c,i}(k+1) = \alpha_{o,i} \cdot \Delta P_{o,i}(k) + P_{u,i}(k) + w_{i,4} \tag{4.56}$$

$$T_{d,i}(k+1) = T_{d,i}(k) + (1-\alpha_{T,i})\left[\eta_i \cdot P_{u,i}(k) - T_{d,i}(k)\right] + w_{i,5}\,. \tag{4.57}$$

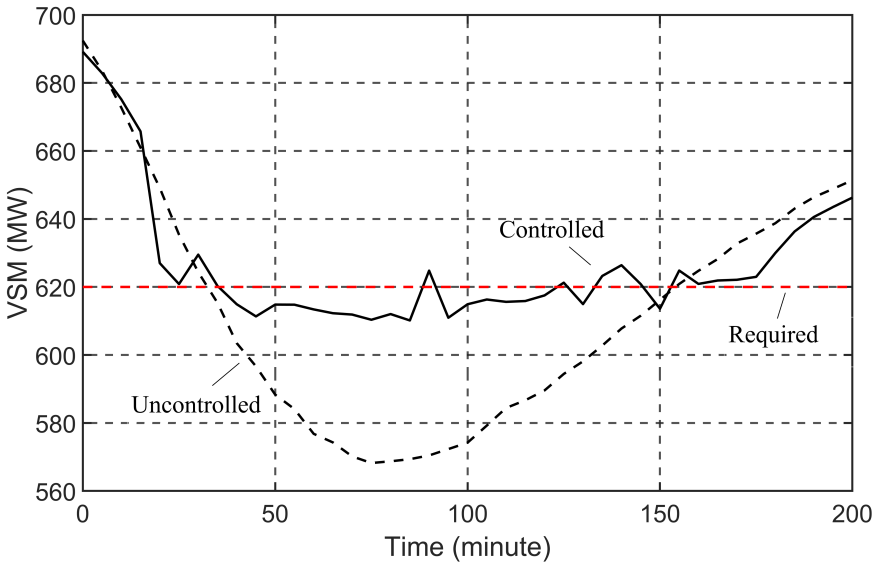


Figure 4.9: Controlled VSM rides through the peak load hours in Scenario 1.

when they are used in the simulation step 2 listed above, where $w_{i,1} \sim w_{i,5}$ are random disturbances. They represent the differences between the simulation model and the real system.

4.6.3.1 Scenario 1: peak load hours

The controlled VSM in Scenario 1 is shown in Figure 4.9. The controller achieved the goal of maintaining VSM, and the valley of the uncontrolled VSM profile is shifted. Due to the errors of linearization and VSM monitoring tool, the controlled VSM is under the limit for some while (note: this is not a steady bias; the error could be either positive or negative).

The control action (load change reference) and the corresponding controlled total load are shown in Figure 4.10. From Figure 4.10(b) we can see the control shifted the load peak. This is different from the peak load shifting service widely implemented in the DR programs. Here the load is shifted as a result of VSM maintenance.

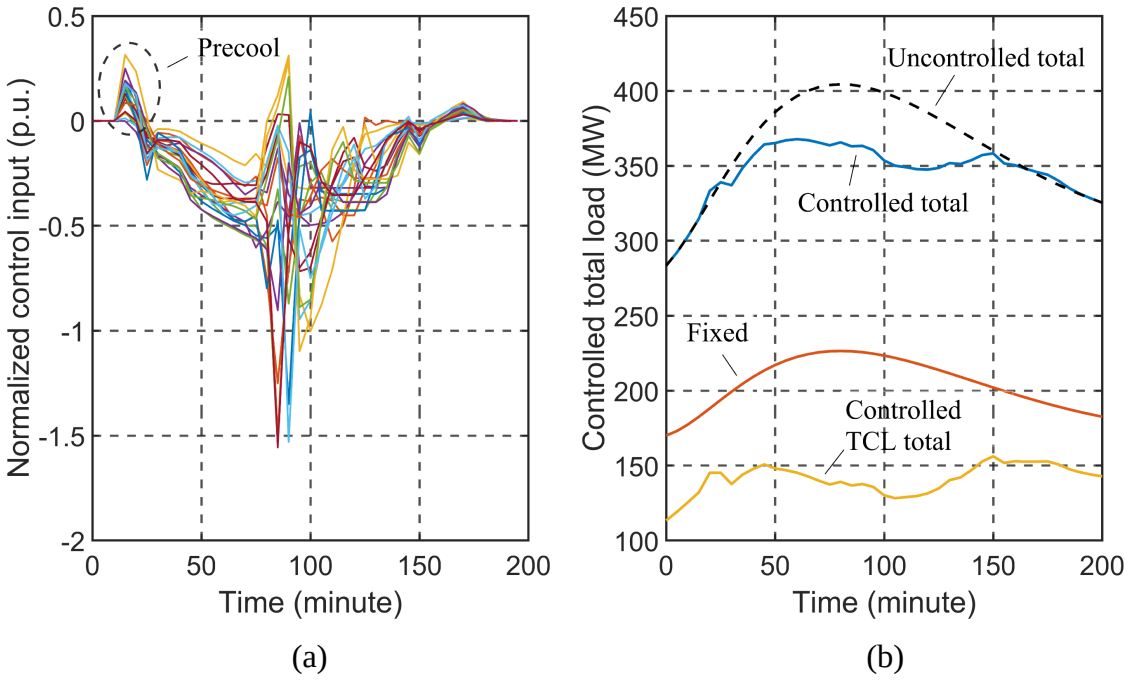


Figure 4.10: Control actions and total load in Scenario 1.(a) Normalized control actions ($P_{u,i}$). Each $P_{u,i}$ is normalized by $P_{o,i,\text{initial}}$. (b) Total load and its two components. The blue curve is the summation of the orange and the yellow. The uncontrolled total load is also plotted for contrast.

The pre-cool feature (see section 4.3.2), based on the look-ahead capability of MPC, can be observed in Figure 4.10(b) (see the dashed circle). Notice that VSM does not violate the limit until $t = 32.65$ (see Figure 4.7). Thus, without a look-ahead framework, there will be no control action before that moment. The proposed controller, via looking ahead, predicted the VSM *will* violate the limit, and acted in advance. The control actions pre-cooled the rooms to minimize the overall control cost within the prediction horizon (defined by (4.52)). This can be clearly seen in Figure 4.11. The in-door temperature deviation is first negative and then positive, thus reduces the magnitude of customer dissatisfaction.

From Figure 4.11 we can see, with the load reduction, the in-door temperatures deviation rose as high as 3.5°C. They caused the customer dissatisfaction increased, but some of them were capped

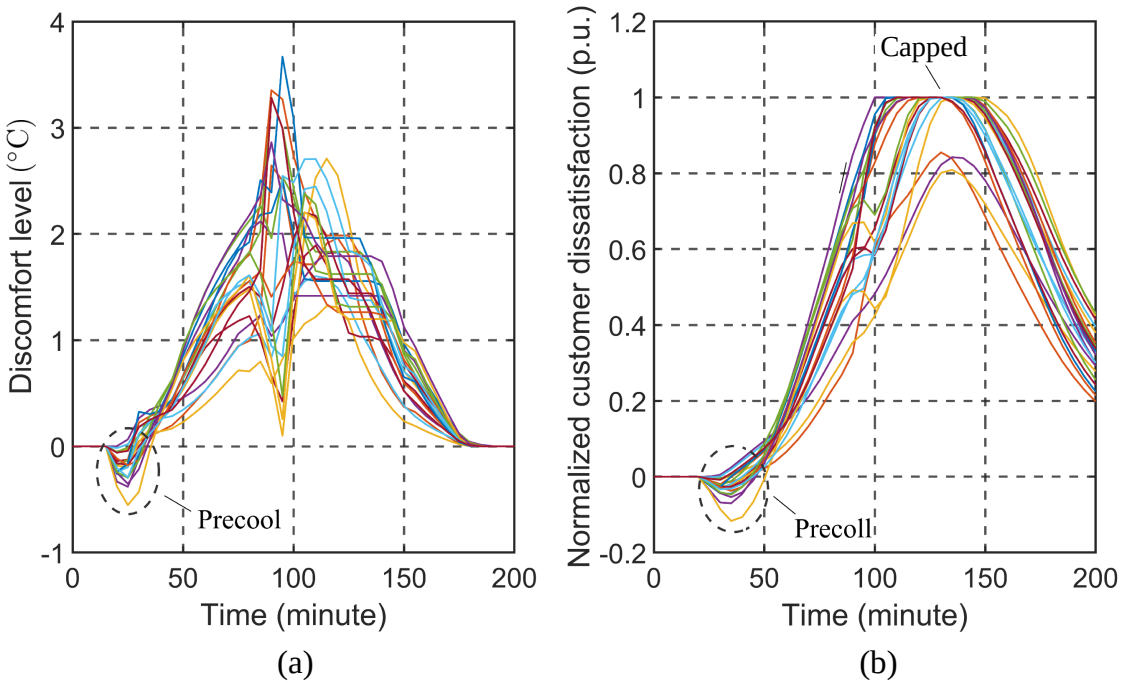


Figure 4.11: Aggregated customer discomfort level and dissatisfaction in Scenario 1. (a) Aggregated customer discomfort level (aggregated in-door temperature deviation from the preference. (b) Normalized aggregated customer dissatisfaction. Each R_i is normalized by $R_{i,\max}$.

(saturated) at the limits. The limits force the control to use other unsaturated TCLs. It reflects the cooperation between control center and customers, and the cooperation among the customers.

4.6.3.2 Scenario 1 with random disturbances

We tested the control performance in Scenario 1 under random disturbances. The disturbances are sampled from independent Gaussian distributions with zero mean. Regarding the standard deviation, we set $2\sigma[w_{i,1}] = \%5R_{i,\max}$, $2\sigma[w_{i,2}] = \%5P_{f,i}(0)$, $2\sigma[w_{i,3}] = 2\sigma[w_{i,4}] = \%5P_{f,o}(0)$, $2\sigma[w_{i,5} = 0.2$, where $\sigma[\cdot]$ denote the standard deviation. That is to say, the disturbances are symmetrically bounded by those specific numbers with probability 0.95, roughly. %5 error is a reasonable assumption for the DR related near-real-time load forecasting, based on the number reported in [154]. The disturbances reflect the discrepancy between the predictive model in MPC and the reality, which is inevitable in practice.

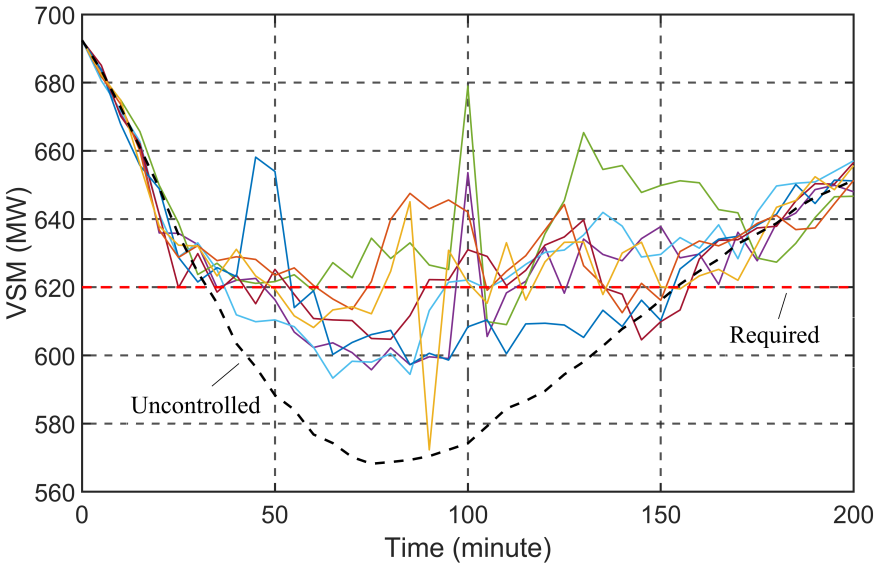


Figure 4.12: Controlled VSM in Scenario 1 under random disturbances. 10 instances are presented.

The controlled VSM under disturbances are shown in Figure 4.12. 10 instances are presented. During the peak time, they are roughly lie in a ± 20 MW band surround the VSM limit. Some extremely spike can be prevented by a simple checking logic on the control action, which will not be concerned in this work. This observation implies that a secure margin for the VSM limit M_r must be considered in practice to incorporate the uncertain disturbance of the system. Estimation of the error band width based on robust optimization techniques will be considered in future.

Besides, we observed that if the robust regularization for α_x, α_u (introduced in section 4.2.3) is not used, the simulation for the random disturbed case fails very often (control action makes infeasible OPF problem). It implies the necessity of the robust correction strategy.

4.6.3.3 Scenario 2: peak load hours + contingency

The controlled VSM in Scenario 2 is shown in Figure 4.13. The controller successfully maintained VSM under the contingency occurs in the peak load time. The "fast cooling" process can be observed in this case. After the contingency was recovered at $t = 150$, VSM rose to roughly 620 MW (above the limit) while the loads were decreasing. Then the controller reversed the control inputs for the aggregators from load reduction to load increase (Figure 4.14), i.e., turning on the

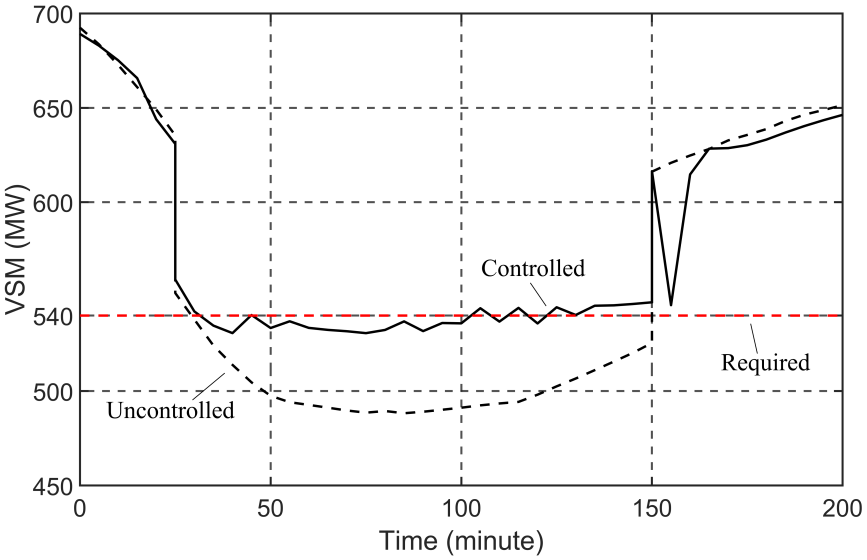


Figure 4.13: Controlled VSM in Scenario 2.

TCLs or decreasing their temperature set-point, which drew VSM back to the level slightly above the limit. This load increase will quickly cool down the rooms, which relieves the discomfort of the customers, and quickly releases the dissatisfaction accumulated during the contingency (see Figure 4.15). The fast cooling process takes only one control step to relief the discomfort levels (see Figure 4.15(a)).

The fast cooling process also shows the necessity of including T_d in the objective in addition to P_u. Without T_d in the objective, the optimal control input P_u must be zero after the contingency was recovered, since there was no constraint violated. In this situation, customers suffered unnecessary discomfort. Besides, since the customer dissatisfaction was quickly released, the DR aggregators now have more reserved capability if another contingency happens soon.

4.6.4 Simulation Platform and Computational Considerations

The proposed controller is implemented on MATLAB. YALMIP [155] is used to formulate the MPC problem as a standard quadratic programming and call Gurobi to solve. MatPower and TSPOPF [156] is used in solving the economic dispatch problem and evaluating the sensitivity

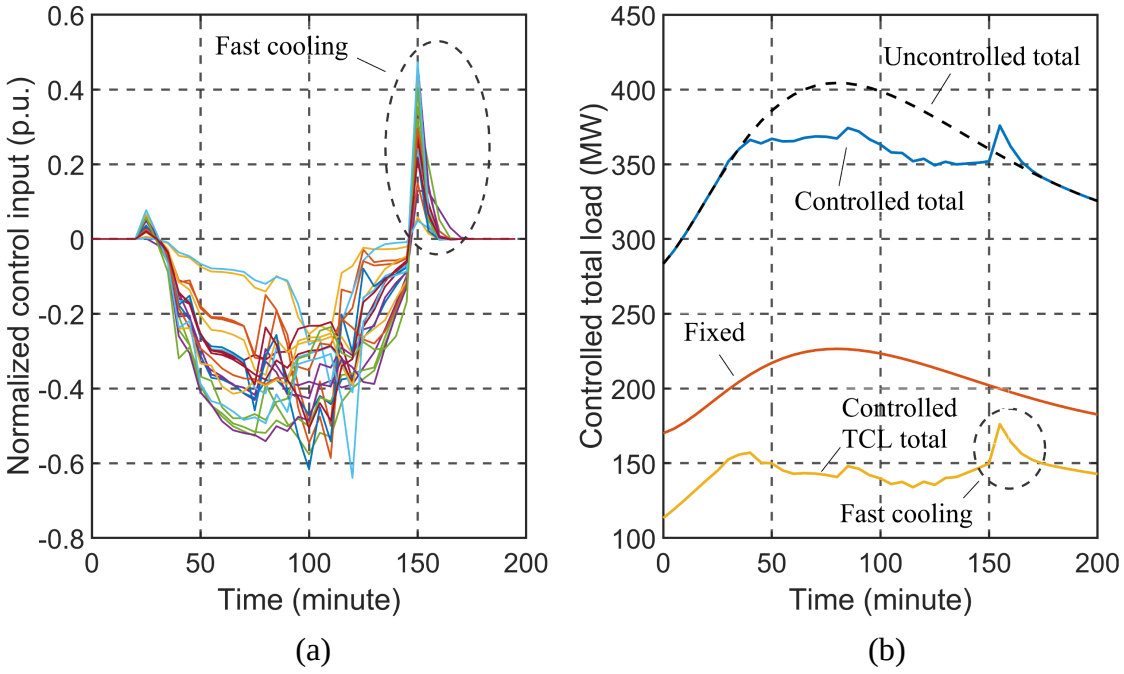


Figure 4.14: Control actions and total load in Scenario 2.(a) Normalized control actions ($P_{u,i}$). Each $P_{u,i}$ is normalized by $P_{o,i,\text{initial}}$. (b) Total load and its two components. The blue curve is the summation of the orange and the yellow. The uncontrolled total load is also plotted for contrast.

matrices. PSS/E is used to trace PV curves. Python is used to manipulate PSS/E and interface PSS/E to Matlab.

A personal computer with 4 cores, 2.9 GHz CPU and 32 GB memory is used to simulate the control in the test example. Solving the MPC takes less than 0.5 second in the test. Several seconds may needed for one control step if several correction steps are involved. For the near-real-time application, solving the MPC is not a very computationally challenging problem even for a practical size system. [157] shows that a problem with state dimension n, control input dimension m, and prediction horizon H takes $O(H(n+m)^3)$ operations per step in an interior-point method for MPC. In our problem, there exists the opportunity to significantly reduce the computationally bounden by adopting distributed MPC techniques, since the states of difference aggregators are almost decoupled, which leads to a highly sparse constraint matrix. There are a

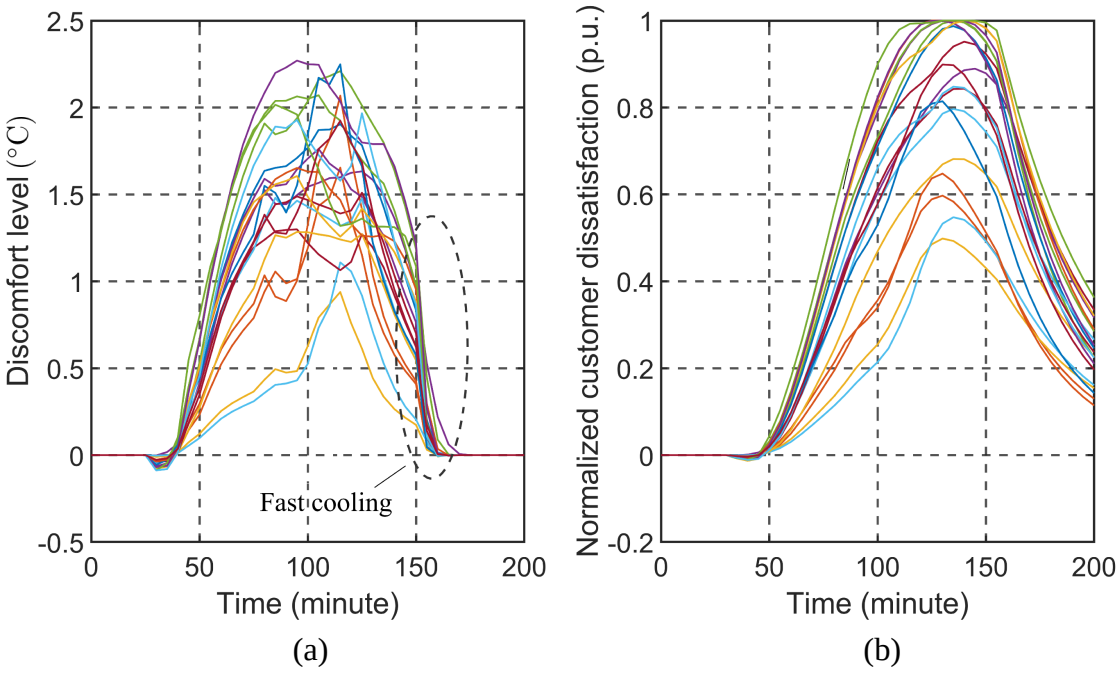


Figure 4.15: Aggregated customer discomfort level and dissatisfaction in Scenario 2. (a) Aggregated customer discomfort level (aggregated in-door temperature deviation from the preference. (b) Normalized aggregated customer dissatisfaction. Each R_i is normalized by $R_{i,\max}$.

number of commercial tools specifically for MPC implementation, which have been widely practiced in industry [158, 159].

4.7 Conclusion

A novel linear MPC-based control approach for maintaining voltage stability margin (VSM) in near-real-time is proposed in this work. It utilizes the thermostatically controlled loads (TCLs) through the demand response aggregators (DR aggregators) as the control measure, thus significantly reduces the cost of demand side stability control traditionally based on under-voltage/frequency load shedding. The cumulative discomfort level of customers due to DR is defined as the customer dissatisfaction and explicitly constrained in the MPC. Thus, the actual influence of DR to the customers is controlled, and the customers are engaged in DR cooperatively. The proposed controller, running online in a receding horizon manner, maintains VSM while minimizing the overall

discomfort level and the control action (mainly the load reduction) within the prediction horizon. The linear predictive model of VSM, which can be provided by the VSM monitoring tool, is utilized to formulate an explicit VSM constraint in the MPC. An adaptive and robust strategy is proposed to iteratively correct the VSM predictive model before the optimal control action is applied, which effectively reduces the control error caused by the discrepancy between the predictive model of MPC and the reality. The effectiveness of the proposed control approach is demonstrated on IEEE 30-bus system, considering the applications in peak load hours and contingency.

CHAPTER 5. FINAL CONCLUSIONS

5.1 Final Conclusions

A comprehensive methodology of online long-term voltage stability monitoring and control is developed for maintaining power system stability and security. New resources for voltage stability analysis and control, namely the knowledge implicit in the online data, and the load flexibility provided by demand response, are exploited to provide timely prediction and control of voltage stability margin (VSM) in an uncertain and non-stationary operating environment.

For real-time VSM monitoring, a machine learning based adaptive approach is proposed. LASSO is tailored to establish an online local regression method that learns the statistical relationship between VSM and online measurements including reactive power reserves (RPRs). This relationship, in form of a local linear predictive model, then can be utilized online to predict VSM as well as its prediction interval, following the real-time measurements collected by SCADA/EMS in a control center. A database updating strategy is proposed to further enhance the adaptivity of the proposed local predictive model. Simulations on several test systems, including a real large-scale system model from industry, showcases the capability of the proposed VSM monitoring tool in tracking the VSM under non-stationary operating condition.

For near-real-time VSM control, a model predictive control (MPC) approach is proposed. Demand response (DR) of thermostatically controlled loads is utilized through a DR aggregator in maintaining VSM under emergency. It significantly reduces the cost of demand side stability control traditionally based on under-voltage/frequency load shedding. Customer dissatisfaction is defined as the cumulative customer discomfort level. It measures the real impact of the demand control on DR participants. Using the model parameters provided by the DR aggregators, the aggregated customer dissatisfaction is explicitly constrained in the MPC to follow the concept of cooperative control. The local linear predictive model of VSM is utilized to formulate an explicit and convex

VSM constraint that is originally implicit, nonconvex, and even discontinuous. It can be obtained from the VSM monitoring tool through a modified procedure, and iteratively corrected through an adaptive and robust strategy. Without violating the customer dissatisfaction constraint, the MPC seeks the optimal control input that maintains VSM in the secure range, while minimizing the overall customer discomfort level and the control actions within the prediction horizon. The performance of the proposed control in peak load hours and contingency is demonstrated on IEEE 30-bus system.

The proactive, adaptive, and cooperative design of the developed approach extends the existing researches on online voltage stability monitoring and control. The methodology developed in this work for VSM, indeed, is general enough to be extended to other long-term stability or security indices, providing a versatile framework that supports the transmission operators in maintaining power system stability and security.

5.2 Research Contributions

5.2.1 Adaptive Real-Time VSM Monitoring Approach

1. Established a framework for real-time VSM monitoring that integrates off-line and online information, via the adaptive regression algorithm (local linear regression) and the adaptive database. It can provide timely VSM prediction on the changing operating condition.
2. Provided a time-varying estimation of the prediction interval along with the VSM prediction, which enhances operators' awareness about how they can trust the VSM prediction and where the true value of VSM could locate, then the closed-loop corrective adaptation can be established (bad prediction can be automatically rectified).
3. Combined local linear regression and LASSO via the relative regularization factor, so as to achieve sufficient scalability for large-scale power systems.

5.2.2 MPC And DR Based VSM Enhancement Approach

1. Established a framework for near-real-time VSM enhancement using linear MPC. It maintains VSM within security limit while minimizing the overall customer discomfort level and control actions within the prediction horizon.
2. Utilized the VSM predictive model that can be obtained from the proposed VSM monitoring tool to formulate the local linear VSM constraint. Provided the modification strategies, namely shrinking the effective domain of control and equally weighting the neighbors, to make the VSM predictive model suitable for control.
3. Provided an adaptive and robust method to iteratively correct the VSM predictive model in one control step until reliable control actions are obtained, leveraging the online nonlinear system model and VSM monitoring tool.
4. Defined customer dissatisfaction as the cumulative customer discomfort level to measure the real impact of the demand control on DR participants. Then explicitly constrained the customer dissatisfaction in the control optimization according to customers' intentions.

5.3 Future Research

5.3.1 Predicting Negative VSM for Diverged What-if Scenarios

Online security assessment usually considers not only the current operating point, but also a credible set of what-if scenarios (e.g. N−1 contingencies). The proposed VSM monitoring tool can provide the VSM prediction for a what-if scenario once the power flow solution is provided. However, some of the what-if conditions may lead to diverged power flow problems. Unfortunately, it is these scenarios that are concerned most by the operators. Thus, people expect the VSM monitoring tool can also give negative VSM for these cases to tell how much load reduction is necessary to achieve an operational state.

Solving the VSM constrained control problem obviously can get such a negative VSM, but it is time consuming and unsuitable for fast online contingency screening. In principle, the machine

learning framework proposed in this work is general enough to be used for negative VSM cases once efficient input features and the training data can be obtained. This is supposed to be a promising research problem.

5.3.2 Forgetting Mechanism for The Database

Notice that the proposed VSM monitoring approach requires the database stays in the memory of the computer. So database cannot be augmented infinitely, and we must provide a strategy, the *forgetting mechanism*, to clean out ineffective data from the database and release the memory. Moreover, from a long-term (i.e. years) point of view, the system is evolving: load grows; new components are commissioned while old ones retire; the network keeps expanding and the topology is changing. Even within a year, power system operating condition changes seasonally. Thus, cleaning out the outdated data is actually a way to achieve long-term data adaptivity.

The weights of data points used in local regression provide a pathway to fulfill this goal. The forgetting effect, reflecting the age (passive, or time-weighted forgetting) or the importance (active, or error-based forgetting) of data [139], can be modeled by a (or a few) dimension(s) of the neighborhood space, thus, old or rarely-accessed data tend to be far away from the current operating point. Alternatively, it can be modeled by one separate forgetting factor that modulates the weights given by KNN and tri-cubic kernel.

The major challenges for realizing forgetting mechanism include: (i) some critical operating conditions, such as the severe contingencies used in off-line initial database generation, may never be accessed and get older as time goes by, but the forgetting mechanism should preventively preserve them in the database (some kind of long-term memory is necessary); (ii) it is difficult to systematically test and verify the forgetting mechanism. Resolving these issues could yield an effective forgetting mechanism that further improve the applicability of the VSM monitoring tool.

5.3.3 Integrating Deep Learning for VSM Prediction

Besides the adaptive approach established in Chapter 3, we are also looking for the opportunity of using deep learning to build a purely off-line model with better generalization performance as the alternative for the multi-linear regression method, and finally move towards a hybrid framework to take advantage of both adaptive model and off-line deep model. We have initialized the application of convolutional neural network for this purpose in [160].

5.3.4 Upgrading to Robust MPC

Model inaccuracy, as the random disturbance, is inevitable in the predictive model. Although this inaccuracy is somewhat considered in designing and testing the control approach, it is not explicated formulated in the control optimization. Robust MPC is supposed to be the proper framework to explicitly incorporate this uncertainty. The challenge is, the computational burden for such methods is usually unaffordable for online application. For example, in our preliminary study, the constraint matrix based on a published closed-loop robust MPC design for IEEE 30-bus system could have more than 1,000,000 columns. However, due to weak coupling among the aggregators, there is potential opportunity to utilize the problem structure to significantly reduce the dimension. Distributed MPC techniques for this purpose could be a profitable research direction.

5.3.5 Integrating Reinforcement Learning for VSM Control

Reinforcement learning is also a very active research field in machine learning which inherently connects to model predictive control [161, 162]. Some efforts have been made for integrating the two methodologies [163]. We are interested in exploring the possibility of involving reinforcement learning in the VSM control problem, either for simply speeding up the solution process of MPC, or for establishing a hybrid framework leveraging both model and data.

5.4 Publications

Journal

Li, Shiyang, and Venkataramana Ajjarapu. "Adaptive Online Monitoring of Voltage Stability Margin via Local Regression." *Power Systems, IEEE Transactions on*, 2017.

Conferences

Li, Shiyang, and Venkataramana Ajjarapu. "Real-time monitoring of long-term voltage stability via local linear regression." *Power & Energy Society General Meeting, 2015 IEEE.*

Li, Shiyang, and Venkataramana Ajjarapu. "Real-Time Monitoring of Long-Term Voltage Stability via Convolutional Neural Network." *Power & Energy Society General Meeting, 2017 IEEE.*

BIBLIOGRAPHY

[1] Y. Mansour *et al.*, "Voltage stability of power systems: concepts, analytical tools, and industry experience," Tech. Rep., 1990.

[2] N. I. D. T. Force, *Survey of the Voltage Collapse Phenomenon.* North American Electric Reliability Council, aug 1991.

[3] C. Canizares, I. Dobson, N. Miller, V. Ajjarapu, H. Hamadanizadeh, and F. Rahimi, "Voltage stability assessment: concepts, practices and tools," Power System Dynamic Performance Committee/Power System Stability Subcommittee, IEEE/PES Power System Stability Subcommittee Special Publication PES-TR9 (Formerly SP101PSS), August 2002.

[4] P. Kundur, J. Paserba, V. Ajjarapu, G. Andersson, A. Bose, C. Canizares, N. Hatziargyriou, D. Hill, A. Stankovic, C. Taylor, T. V. Cutsem, and V. Vittal, "Definition and classification of power system stability IEEE/cigre joint task force on stability terms and definitions," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1387–1401, Aug. 2004.

[5] C. W. Taylor, *Power system voltage stability.* McGraw-Hill, 1994.

[6] V. Ajjarapu, *Computational techniques for voltage stability assessment and control.* Springer Science+Business Media, LLC, 2006.

[7] "Final report on the august 14, 2003 blackout in the united states and canada: causes and recommendations," U.S.-Canada Power System Outage Task Force, Tech. Rep., April 2014.

[8] V. A. Venikov, V. A. Stroev, V. I. Idelchick, and V. I. Tarasov, "Estimation of electrical power system steady-state stability in load flow calculations," *IEEE Trans. Power App. Syst.*, vol. 94, no. 3, pp. 1034–1041, May 1975.

[9] H. Kwatny, A. Pasrija, and L. Bahar, "Static bifurcations in electric power networks: Loss of steady-state stability and voltage collapse," *IEEE Trans. Circuits Syst.*, vol. 33, no. 10, pp. 981–991, Oct. 1986.

[10] P. W. Sauer and M. A. Pai, "Power system steady-state stability and the load-flow jacobian," *IEEE Trans. Power Syst.*, vol. 5, no. 4, pp. 1374–1383, Nov. 1990.

[11] I. Dobson and L. Lu, "Voltage collapse precipitated by the immediate change in stability when generator reactive power limits are encountered," *IEEE Trans. Circuits Syst. I*, vol. 39, no. 9, pp. 762–766, Sep. 1992.

[12] T. Van Cutsem and C. Vournas, *Voltage stability of electric power systems*. Springer US, 1998, vol. 441.

[13] A. J. Korsak, "On the question of uniqueness of stable load-flow solutions," *IEEE Trans. Power App. Syst.*, vol. PAS-91, no. 3, pp. 1093–1100, May 1972.

[14] S. S. Sastry, *Nonlinear systems: analysis, stability, and control*. Springer Science & Business Media, 2013, vol. 10.

[15] A. M. Chebbo, M. R. Irving, and M. J. H. Sterling, "Voltage collapse proximity indicator: behaviour and implications," *Transmission and Distribution IEE Proceedings C-Generation*, vol. 139, no. 3, pp. 241–252, May 1992.

[16] C. Canizares, A. De Souza, and V. Quintana, "Comparison of performance indices for detection of proximity to voltage collapse," *IEEE Trans. Power Syst.*, vol. 11, no. 3, pp. 1441–1450, 1996.

[17] M. M. El-Kateb, S. Abdelkader, and M. S. Kandil, "Linear indicator for voltage collapse in power systems," *Transmission and Distribution IEE Proceedings-Generation*, vol. 144, no. 2, pp. 139–146, Mar. 1997.

[18] C. Reis and F. P. M. Barbosa, "Indicators for voltage collapse margins," in *Proc. Asia-Pacific Power and Energy Engineering Conf*, Mar. 2010, pp. 1–4.

[19] M. Cupelli, C. Cardet, and A. Monti, "Comparison of line voltage stability indices using dynamic real time simulation," in *Innovative Smart Grid Technologies (ISGT Europe), 2012 3rd IEEE PES International Conference and Exhibition on*, 2012, pp. 1–8.

[20] B. Leonardi and V. Ajjarapu, "Investigation of various generator reactive power reserve (grpr) definitions for online voltage stability/security assessment," in *Proc. IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, Jul. 2008, pp. 1–7.

[21] *Monitoring System Conditions*, North American Electric Reliability Corporation (NERC) Std. TOP-006-3, November 2012.

[22] *Voltage and Reactive Control*, North American Electric Reliability Corporation (NERC) Std. VAR-001-4.1, nov 2015.

[23] F. E. R. Commission *et al.*, "Frequency regulation compensation in the organized wholesale power markets," *Order*, no. 755, p. 76, 2011.

[24] Department of Energy. Demand response. [Online]. Available: https://energy.gov/oe/activities/technology-development/grid-modernization-and-smart-grid/demand-response

[25] D. J. Olsen, N. Matson, M. D. Sohn, C. Rose, J. Dudley, S. Goli, S. Kiliccote, M. Hummon, D. Palchak, P. Denholm *et al.*, "Grid integration of aggregated demand response, part 1: load availability profiles and constraints for the western interconnection," Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA (United States); National Renewable Energy Laboratory (NREL), Golden, CO (United States), Tech. Rep., 2013.

[26] *System Operating Limits Methodology for the Operations Horizon*, North American Electric Reliability Corporation (NERC) Std. TOP-006-3, November 2015.

[27] F. Dong, B. H. Chowdhury, M. L. Crow, and L. Acar, "Improving voltage stability by reactive power reserve management," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 338–345, Feb. 2005.

[28] Q. Sun, H. Cheng, D. Yang, S. Fang, J. Zhang, B. Li, and Y. Song, "A method to improve reactive reserve management with respect to voltage stability," in *Proc. IEEE Power Energy Society General Meeting*, Jul. 2015, pp. 1–6.

[29] FERC, "Federal energy regulatory commission staff preliminary assessment of the north american electric reliability councils proposed mandatory reliability standards," FERC, Tech. Rep. RM06-16-000, 2006.

[30] K. Iba, H. Suzuki, M. Egawa, and T. Watanabe, "Calculation of critical loading condition with nose curve using homotopy continuation method," *IEEE Trans. Power Syst.*, vol. 6, no. 2, pp. 584–593, May 1991.

[31] V. Ajjarapu and C. Christy, "The continuation power flow: a tool for steady state voltage stability analysis," *IEEE Trans. Power Syst.*, vol. 7, no. 1, pp. 416–423, 1992.

[32] H.-D. Chiang, A. J. Flueck, K. S. Shah, and N. Balu, "CPflow: a practical tool for tracing power system steady-state stationary behavior due to load and generation variations," *IEEE Trans. Power Syst.*, vol. 10, no. 2, pp. 623–634, May 1995.

[33] R. Seydel, "Numerical computation of branch points in nonlinear equations," *Numerische Mathematik*, vol. 33, no. 3, pp. 339–352, 1979.

[34] V. Ajjarapu, "Identification of steady-state voltage stability in power systems," *International Journal of Energy Systems*, vol. 11, no. 1, pp. 43–46, 1991.

[35] H.-D. Chiang and R. Jean-Jumeau, "A more efficient formulation for computation of the maximum loading points in electric power systems," *IEEE Trans. Power Syst.*, vol. 10, no. 2, pp. 635–646, May 1995.

[36] G. D. Irisarri, X. Wang, J. Tong, and S. Mokhtari, "Maximum loadability of power systems using interior point nonlinear optimization method," *IEEE Trans. Power Syst.*, vol. 12, no. 1, pp. 162–172, Feb. 1997.

[37] L. A. L. Zarate, C. A. Castro, J. L. M. Ramos, and E. R. Ramos, "Fast computation of voltage stability security margins using nonlinear programming techniques," *IEEE Trans. Power Syst.*, vol. 21, no. 1, pp. 19–27, Feb. 2006.

[38] M. Glavic and T. Van Cutsem, "A short survey of methods for voltage instability detection," in *Power and Energy Society General Meeting, 2011 IEEE*, 2011, pp. 1–8.

[39] H.-D. Chiang and R. Jean-Jumeau, "Toward a practical performance index for predicting voltage collapse in electric power systems," *IEEE Trans. Power Syst.*, vol. 10, no. 2, pp. 584–592, May 1995.

[40] L. Warland and A. T. Holen, "Estimation of distance to voltage collapse: Testing an algorithm based on local measurements," *14th PSCC Sevilla*, 2002.

[41] A. R. R. Matavalam and V. Ajjarapu, "Calculating the long term voltage stability margin using a linear index," in *Proc. IEEE Power Energy Society General Meeting*, Jul. 2015, pp. 1–5.

[42] Z. Wang, B. Cui, and J. Wang, "A necessary condition for power flow insolvability in power distribution systems with distributed generators," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1440–1450, Mar. 2017.

[43] L. Xie, Y. Chen, and H. Liao, "Distributed online monitoring of quasi-static voltage collapse in multi-area power systems," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 2271–2279, Nov. 2012.

[44] J. W. Simpson-Porco and F. Bullo, "Distributed monitoring of voltage collapse sensitivity indices," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 1979–1988, Jul. 2016.

[45] X. Liu, X. Zhang, and V. Venkatasubramanian, "Distributed voltage security monitoring in large power systems using synchrophasors," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 982–991, Mar. 2016.

[46] K. Vu, M. Begovic, D. Novosel, and M. Saha, "Use of local measurements to estimate voltage-stability margin," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 1029–1035, 1999.

[47] B. Genêt and J.-C. Maun, "Voltage-stability monitoring using wide-area measurement systems," in *Power Tech, 2007 IEEE Lausanne*. IEEE, 2007, pp. 1712–1717.

[48] S. Corsi and G. Taranto, "A real-time voltage instability identification algorithm based on local phasor measurements," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1271–1279, 2008.

[49] M. Glavic and T. Van Cutsem, "Wide-area detection of voltage instability from synchronized phasor measurements. part i: Principle," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1408–1416, 2009.

[50] Y. Tamura, H. Mori, and S. Iwamoto, "Relationship between voltage instability and multiple load flow solutions in electric power systems," *IEEE Trans. Power App. Syst.*, vol. PAS-102, no. 5, pp. 1115–1125, May 1983.

[51] Y. Tamura, K. Sakamoto, and Y. Tayama, "Voltage instability proximity index (vipi) based on multiple load flow solutions in ill-conditioned power systems," in *Proc. 27th IEEE Conf. Decision and Control*, Dec. 1988, pp. 2114–2119 vol.3.

[52] C. L. DeMarco and T. J. Overbye, "An energy based security measure for assessing vulnerability to voltage collapse," *IEEE Trans. Power Syst.*, vol. 5, no. 2, pp. 419–427, May 1990.

[53] F. Gubina and B. Strmcnik, "Voltage collapse proximity index determination using voltage phasors approach," *IEEE Trans. Power Syst.*, vol. 10, no. 2, pp. 788–794, May 1995.

[54] G. Verbic and F. Gubina, "A new concept of voltage-collapse protection based on local phasors," *IEEE Trans. Power Del.*, vol. 19, no. 2, pp. 576–581, Apr. 2004.

[55] I. Smon, G. Verbic, and F. Gubina, "Local voltage-stability index using tellegen's theorem," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1267–1275, 2006.

[56] M. Haque, "Use of local information to determine the distance to voltage collapse," in *Power Engineering Conference, 2007. IPEC 2007. International.* IEEE, 2007, pp. 407–412.

[57] S. M. Abdelkader and D. J. Morrow, "Online th évenin equivalent determination considering system side changes and measurement errors," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2716–2725, Sep. 2015.

[58] Y. Wang, I. R. Pordanjani, W. Li, W. Xu, T. Chen, E. Vaahedi, and J. Gurney, "Voltage stability monitoring based on the concept of coupled single-port circuit," *Power Systems, IEEE Transactions on*, vol. 26, no. 4, pp. 2154–2163, 2011.

[59] H. Yuan and F. Li, "A comparative study of measurement-based thevenin equivalents identification methods," in *Proc. North American Power Symp. (NAPS)*, Sep. 2014, pp. 1–6.

[60] F. Hu, K. Sun, A. D. Rosso, E. Farantatos, and N. Bhatt, "Measurement-based real-time voltage stability monitoring for load areas," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2787–2798, Jul. 2016.

[61] H. Y. Su and C. W. Liu, "Estimating the voltage stability margin using pmu measurements," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3221–3229, Jul. 2016.

[62] H. Yuan, X. Li, F. Li, X. Fang, H. Cui, and Q. Hu, "Mitigate overestimation of voltage stability margin by coupled single-port circuit models," in *Proc. IEEE Power and Energy Society General Meeting (PESGM)*, Jul. 2016, pp. 1–5.

[63] R.-B. Li, Z.-Q. Wu, J.-W. Ge, D. Liu, W.-H. Zhu, Y.-H. Huang, and C. Zhang, "Analysis for static-state voltage stability based on obtaining thevenin equivalent parameters with

sensitivity method," in *Proceedings of the CSU-EPSA*, vol. 23, no. 6. Editorial Board of Proceedings of the CSU-EPSA, School of Electrical & Automation Engineering Tianjin 300072 China, 2011, pp. 76–80.

[64] A. R. R. Matavalam and V. Ajjarapu, "Sensitivity based thevenin index with systematic inclusion of reactive limits," *IEEE Trans. Power Syst.*, 2017.

[65] L. Sandberg and K. Roudén, "The real-time supervision of transmission capacity in the swedish grid," *Real-time Stability Assessment in Modern Power System Control Centers*, pp. 279–305, 2008.

[66] H.-D. Chiang and P. Causgrove, "Study of co-optimimization stochastic superopf application in the caiso system," Bigwood Systems, Inc., Tech. Rep., sep 2015.

[67] P. Dimo, "Étude de la stabilité statique et du réglage de tension," *RGE, Paris*, vol. 70, no. 11, pp. 552–556, 1961.

[68] ——, *Nodal analysis of power systems.* International Scholarly Book Services, Inc., Forest Grove, OR, 1975.

[69] S. C. Savulescu, M. L. Oatts, J. G. Pruitt, F. Williamson, and R. Adapa, "Fast steady-state stability assessment for real-time and operations planning," *IEEE Trans. Power Syst.*, vol. 8, no. 4, pp. 1557–1569, Nov. 1993.

[70] S. C. Savulescu, "Fast assessment of the distance to instability," *Real-Time Stability in Power Systems: Techniques for Early Detection of the Risk of Blackout*, p. 31, 2006.

[71] F. Wu and S. Kumagai, "Steady-state security regions of power systems," *IEEE Trans. Circuits Syst.*, vol. 29, no. 11, pp. 703–711, Nov. 1982.

[72] Y. Yu, "Security region of bulk power system," in *Proc. Int. Conf. Power System Technology*, vol. 1, Oct. 2002, pp. 13–17 vol.1.

[73] J. Zhu, *Steady-State Security Regions.* Wiley-IEEE Press, 2015, pp. 664–.

[74] S. Chen, Z. Wei, G. Sun, Y. Sun, and N. Lu, "Steady-state security regions of electricity-gas integrated energy systems," in *Proc. IEEE Power and Energy Society General Meeting (PESGM)*, Jul. 2016, pp. 1–5.

[75] J. Su, Y. Yu, H. Jia, P. Li, N. He, Z. Tang, and H. Fu, "Visualization of voltage stability region of bulk power system," in *Proc. Int. Conf. Power System Technology*, vol. 3, 2002, pp. 1665–1668 vol.3.

[76] J. D. McCalley, S. Wang, Q. L. Zhao, G. Z. Zhou, R. T. Treinen, and A. D. Papalexopoulos, "Security boundary visualization for systems operation," *IEEE Trans. Power Syst.*, vol. 12, no. 2, pp. 940–947, May 1997.

[77] X. Gu and C. A. Canizares, "Fast prediction of loadability margins using neural networks to approximate security boundaries of power systems," *Transmission Distribution IET Generation*, vol. 1, no. 3, pp. 466–475, May 2007.

[78] V. J. Gutierrez-Martinez, C. A. Canizares, C. R. Fuerte-Esquivel, A. Pizano-Martinez, and X. Gu, "Neural-network security-boundary constrained optimal power flow," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 63–72, Feb. 2011.

[79] G. Wang, X. Zhang, and S. Mei, "Quadratic approximation analysis of static voltage stability region boundary," in *Proceedings of the CSEE*, vol. 28, no. 19, 2008, pp. 0–35.

[80] M. Perninge and L. Soder, "On the validity of local approximations of the power system loadability surface," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2143–2153, Nov. 2011.

[81] C. Hamon, M. Perninge, and L. Sder, "A stochastic optimal power flow problem with stability constraints —part i: Approximating the stability boundary," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1839–1848, May 2013.

[82] Y. Qiu, H. Wu, Y. Zhou, and Y. Song, "Global parametric polynomial approximation of static voltage stability region boundaries," *IEEE Trans. Power Syst.*, vol. PP, no. 99, p. 1, 2016.

[83] A. Alves da Silva, "Overcoming limitations of nns for on-line DSA," in *Power Engineering Society General Meeting, 2005. IEEE*, 2005, pp. 2653–2660.

[84] A. El-Keib and X. Ma, "Application of artificial neural networks in voltage stability assessment," *IEEE Trans. Power Syst.*, vol. 10, no. 4, pp. 1890–1896, 1995.

[85] B. Jeyasurya, "Artificial neural networks for on-line voltage stability assessment," in *Power Engineering Society Summer Meeting, 2000. IEEE*, vol. 4, 2000, pp. 2014–2018.

[86] D. Popovic, D. Kukolj, and F. Kulic, "Monitoring and assessment of voltage stability margins using artificial neural networks with a reduced input set," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 145, no. 4, pp. 355–362, 1998.

[87] D. Zhou, U. Annakkage, and A. Rajapakse, "Online monitoring of voltage stability margin using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1566–1574, 2010.

[88] H. Razmi, M. Teshnehlab, and H. Shayanfar, "Neural network based on a genetic algorithm for power system loading margin estimation," *IET Generation, Transmission & Distribution*, vol. 6, no. 11, pp. 1153–1163, 2012.

[89] K.-L. Du and M. N. S. Swamy, *Neural Networks and Statistical Learning.* Springer-Verlag London, 2014.

[90] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, second edition ed. Springer Science+Business Media, LLC, 2009.

[91] R. Setiono, W. K. Leow, and J. Zurada, "Extraction of rules from artificial neural networks for nonlinear regression," *IEEE Trans. Neural Netw.*, vol. 13, no. 3, pp. 564–577, 2002.

[92] A. Bahmanyar and A. Karami, "Power system voltage stability monitoring using artificial neural networks with a reduced set of inputs," *International Journal of Electrical Power & Energy Systems*, vol. 58, pp. 246–256, 2014.

[93] T. Van Cutsem, L. Wehenkel, M. Pavella, B. Heilbronn, and M. Goubin, "Decision tree approaches to voltage security assessment," *IEE Proceedings C Generation, Transmission and Distribution*, vol. 140, no. 3, pp. 189–198, 1993.

[94] R. Diao, K. Sun, V. Vittal, R. J. O'Keefe, M. R. Richardson, N. Bhatt, D. Stradford, and S. K. Sarawgi, "Decision tree-based online voltage security assessment using pmu measurements," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 832–839, May 2009.

[95] C. Zheng, V. Malbasa, and M. Kezunovic, "Regression tree for stability margin prediction using synchrophasor measurements," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1978–1987, 2013.

[96] M. Negnevitsky, N. Tomin, V. Kurbatsky, D. Panasetsky, A. Zhukov, and C. Rehtanz, "A random forest-based approach for voltage security monitoring in a power system," in *Proc. IEEE Eindhoven PowerTech*, Jun. 2015, pp. 1–6.

[97] M. Mohammadi and G. B. Gharehpetian, "Application of core vector machines for on-line voltage security assessment using a decisiontree-based feature selection algorithm," *Transmission Distribution IET Generation*, vol. 3, no. 8, pp. 701–712, Aug. 2009.

[98] Y. Fan, S. Liu, L. Qin, H. Li, and H. Qiu, "A novel online estimation scheme for static voltage stability margin based on relationships exploration in a large data set," *IEEE Trans. Power Syst.*, vol. 30, no. 3, pp. 1380–1393, May 2015.

[99] L. Bao, Z. Huang, and W. Xu, "Online voltage stability monitoring using var reserves," *IEEE Trans. Power Syst.*, vol. 18, no. 4, pp. 1461–1469, 2003.

[100] B. Leonardi and V. Ajjarapu, "Development of multilinear regression models for online voltage stability margin estimation," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 374–383, 2011.

[101] H. Song, B. Lee, S.-H. Kwon, and V. Ajjarapu, "Reactive reserve-based contingency constrained optimal power flow (RCcopf) for enhancement of voltage stability margins," *IEEE Trans. Power Syst.*, vol. 18, no. 4, pp. 1538–1546, Nov. 2003.

[102] O. A. Mousavi, M. Bozorg, A. Ahmadi-Khatir, and R. Cherkaoui, "Reactive power reserve management: Preventive countermeasure for improving voltage stability margin," in *Proc. IEEE Power and Energy Society General Meeting*, Jul. 2012, pp. 1–7.

[103] Q. Sun, H. Cheng, and Y. Song, "Bi-objective reactive power reserve optimization to coordinate long-and short-term voltage stability," *IEEE Access*, vol. PP, no. 99, p. 1, 2017.

[104] S. Fang, H. Cheng, G. Xu, Q. Zhou, H. He, and P. Zeng, "Stochastic optimal reactive power reserve dispatch considering voltage control areas," *International Transactions on Electrical Energy Systems*, vol. 27, no. 3, 2017.

[105] R. R. Negenborn, A. G. Beccuti, T. Demiray, S. Leirens, G. Damm, B. De Schutter, and M. Morari, "Supervisory hybrid model predictive control for voltage stability of power networks," in *American Control Conference, 2007. ACC'07.* IEEE, 2007, pp. 5444–5449.

[106] S. M. Mohseni-Bonab and A. Rabiee, "Optimal reactive power dispatch: a review, and a new stochastic voltage stability constrained multi-objective model at the presence of uncertain wind power generation," *Transmission Distribution IET Generation*, vol. 11, no. 4, pp. 815–829, 2017.

[107] E. Vaahedi, Y. Mansour, C. Fuchs, S. Granville, M. D. L. Latore, and H. Hamadanizadeh, "Dynamic security constrained optimal power flow/var planning," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 38–43, Feb. 2001.

[108] W. D. Rosehart, C. A. Canizares, and V. H. Quintana, "Effect of detailed power system models in traditional and voltage-stability-constrained optimal power-flow problems," *IEEE Transactions on Power Systems*, vol. 18, no. 1, pp. 27–35, 2003.

[109] N. Xiong, H. Cheng, L. Yao, Y. Gu, Z. Ma, Z. Zhu, and J. Zhang, "Investigation on limit surfaces in space spanned by generation parameter," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1309–1318, 2010.

[110] J.-q. Zhao, H.-D. CHIANG, and B.-m. ZHANG, "An on-line enhancement control algorithm for static stability in power system," *Proceedings of the Csee*, vol. 8, p. 001, 2005.

[111] D. Wang, H. Jia, C. Wang, N. Lu, M. Fan, Y. Zhou, and Y. Qi, "Voltage stability enhancement using thermostatically controlled appliances as a comfort-constrained virtual generator," *International Transactions on Electrical Energy Systems*, vol. 25, no. 12, pp. 3509–3522, 2015.

[112] S. Greene, I. Dobson, and F. L. Alvarado, "Sensitivity of the loading margin to voltage collapse with respect to arbitrary parameters," *IEEE Trans. Power Syst.*, vol. 12, no. 1, pp. 262–272, Feb. 1997.

[113] B. Leonardi and V. Ajjarapu, "An approach for real time voltage stability margin control via reactive power reserve sensitivities," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 615–625, 2013.

[114] M. Morari and J. H. Lee, "Model predictive control: past, present and future," *Computers & Chemical Engineering*, vol. 23, pp. 667 – 682, 1999.

[115] J. Rawlings, "Tutorial overview of model predictive control," *IEEE Control Syst. Mag.*, vol. 20, no. 3, pp. 38–52, 2000.

[116] E. F. Camacho and C. Bordons, *Model Predictive Control*, second edition ed. Springer-Verlag London, 2007.

[117] Y. Gu and L. Xie, "Early detection and optimal corrective measures of power system insecurity in enhanced look-ahead dispatch," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1297–1307, 2013.

[118] L. Jin, R. Kumar, and N. Elia, "Model predictive control-based real-time power system protection schemes," *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 988–998, 2010.

[119] A. Venkat, I. Hiskens, J. Rawlings, and S. Wright, "Distributed MPC strategies with application to power system automatic generation control," *IEEE Trans. Control Syst. Technol.*, vol. 16, no. 6, pp. 1192–1206, 2008.

[120] M. Glavic and T. Van Cutsem, "Some reflections on model predictive control of transmission voltages," in *Power Symposium, 2006. NAPS 2006. 38th North American*, 2006, pp. 625–632.

[121] G. Valverde and T. Van Cutsem, "Model predictive control of voltages in active distribution networks," *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 2152–2161, 2013.

[122] A. Kassem and A. Abdelaziz, "Reactive power control for voltage stability of standalone hybrid wind–diesel power system based on functional model predictive control," *IET Renewable Power Generation*, vol. 8, no. 8, pp. 887–899, 2014.

[123] S. Deshmukh, B. Natarajan, and A. Pahwa, "State estimation and voltage/var control in distribution network with intermittent measurements," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 200–209, 2014.

[124] M. Larsson, D. J. Hill, and G. Olsson, "Emergency voltage control using search and predictive control," *International Journal of Electrical Power & Energy Systems*, vol. 24, no. 2, pp. 121 – 130, 2002.

[125] J. Wen, Q. Wu, D. Turner, S. Cheng, and J. Fitch, "Optimal coordinated voltage control for power system voltage stability," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 1115–1122, 2004.

[126] I. A. Hiskens and B. Gong, "Voltage stability enhancement via model predictive control of load," in *Proceedings of the Symposium on Bulk Power System Dynamics and Control VI, Cortina dAmpezzo*, 2004.

[127] M. Glavic, M. Hajian, W. Rosehart, and T. Van Cutsem, "Receding-horizon multi-step optimization to correct nonviable or unstable transmission voltages," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1641–1650, 2011.

[128] M. Hajian, W. Rosehart, M. Glavic, H. Zareipour, and T. Van Cutsem, "Linearized power flow equations based predictive control of transmission voltages," in *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, 2013, pp. 2298–2304.

[129] N. Hatziargyriou, T. Van Cutsem, J. Milanović, P. Pourbeik, C. Vournas, O. Vlachokyriakou, P. Kotsampopoulos, R. Ramos, J. Boemer, P. Aristidou *et al.*, "Contribution to bulk system control and stability by distributed energy resources connected at distribution network," IEEE, Tech. Rep., 2017.

[130] S. S. Baghsorkhi and S. P. Suetin, "Embedding ac power flow in the complex plane part i: Modelling and mathematical foundation," *arXiv preprint arXiv:1604.03425*, 2016.

[131] S. Li and V. Ajjarapu, "Real-time monitoring of long-term voltage stability via local linear regression," in *Proc. IEEE Power Energy Society General Meeting*, Jul. 2015, pp. 1–5.

[132] B. Clarke, E. Fokoue, and H. H. Zhang, *Principles and Theory for Data Mining and Machine Learning*. Springer Science+Business Media, LLC, 2009.

[133] M. Sugiyama and M. Kawanabe, *Machine learning in non-stationary environments: Introduction to covariate shift adaptation*. MIT press, 2012.

[134] B. H. Leonardi, "Investigation of novel methodologies using reactive power reserves for online voltage stability margin monitoring and control," Ph.D. dissertation, Iowa State University, 2011.

[135] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[136] T. Hastie, J. Taylor, R. Tibshirani, G. Walther *et al.*, "Forward stagewise regression and the monotone lasso," *Electronic Journal of Statistics*, vol. 1, pp. 1–29, 2007.

[137] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical learning with sparsity*. CRC press, 2015.

[138] L. Xie, Y. Chen, and P. R. Kumar, "Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2784–2794, Nov. 2014.

[139] L. I. Kuncheva and J. S. Sánchez, "Nearest neighbour classifiers for streaming data with delayed labelling," in *Proc. Eighth IEEE Int. Conf. Data Mining*, Dec. 2008, pp. 869–874.

[140] *Load Shedding Plans*, North American Electric Reliability Corporation (NERC) Std. EOP-003-2, nov 2012.

[141] *Automatic Underfrequency Load Shedding*, North American Electric Reliability Corporation (NERC) Std. PRC-006-3, nov 2017.

[142] R. Faranda, A. Pievatolo, and E. Tironi, "Load shedding: A new proposal," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 2086–2093, Nov. 2007.

[143] D. P. Nedic, I. Dobson, D. S. Kirschen, B. A. Carreras, and V. E. Lynch, "Criticality in a cascading failure blackout model," *International Journal of Electrical Power & Energy Systems*, vol. 28, no. 9, pp. 627–633, 2006.

[144] J. H. Eto, "The national cost of power interruptions to electricity customers - a revised update," Lawrence Berkeley National Laboratory, Tech. Rep., Jan. 2017.

[145] Q. Qdr, "Benefits of demand response in electricity markets and recommendations for achieving them," Tech. Rep., 2006.

[146] J. Devlin and R. S. Worldwide, "Residential energy consumption survey," U.S. Department of Energy, Tech. Rep., 2001.

[147] A. Radaideh, U. Vaidya, and V. Ajjarapu, "Sequential set-point control for heterogeneous thermostatically controlled loads through an extended markov chain abstraction," *IEEE Transactions on Smart Grid*, 2017.

[148] P. Xu and R. Yin, "Pre-cooling strategies for demand response in large commercial buildings," Lawrence Berkeley National Laboratory, Tech. Rep., Oct. 2008.

[149] J. Ma, J. Qin, T. Salsbury, and P. Xu, "Demand reduction in building energy systems based on economic model predictive control," *Chemical Engineering Science*, vol. 67, no. 1, pp. 92–100, 2012.

[150] A. Keane, L. F. Ochoa, E. Vittal, C. J. Dent, and G. P. Harrison, "Enhanced utilization of voltage control resources with distributed generation," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 252–260, Feb. 2011.

[151] N. U. Nandasiri, "Aggregator based optimal operation of radial distribution system," Ph.D. dissertation, Wichita State University, 2016.

[152] A. Singhal and V. Ajjarapu, "A framework to utilize ders var resources to support the grid in an integrated t-d system," submitted to PES General Meeting, 2018.

[153] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[154] L. Neusser and L. N. Canha, "Real-time load forecasting for demand side management with only a few days of history available," in *Power Engineering, Energy and Electrical Drives (POWERENG), 2013 Fourth International Conference on.* IEEE, 2013, pp. 911–914.

[155] J. Lofberg, "Yalmip : a toolbox for modeling and optimization in matlab," in *Proc. IEEE Int. Conf. Robotics and Automation (IEEE Cat. No.04CH37508)*, Sep. 2004, pp. 284–289.

[156] H. Wang, C. E. Murillo-Sanchez, R. D. Zimmerman, and R. J. Thomas, "On computational issues of market-based optimal power flow," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1185–1193, Aug. 2007.

[157] Y. Wang and S. Boyd, "Fast model predictive control using online optimization," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 2, pp. 267–278, Mar. 2010.

[158] S. J. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," *Control engineering practice*, vol. 11, no. 7, pp. 733–764, 2003.

[159] H. J. Ferreau, S. Almr, H. Peyrl, J. L. Jerez, and A. Domahidi, "Survey of industrial applications of embedded model predictive control," in *Proc. European Control Conf. (ECC)*, Jun. 2016, p. 601.

[160] S. Li and V. Ajjarapu, "Real-time monitoring of long-term voltage stability via convolutional neural network," in *2017 IEEE Power Energy Society General Meeting*, July 2017 (accepted for publishing).

[161] D. Ernst, M. Glavic, F. Capitanescu, and L. Wehenkel, "Model predictive control and reinforcement learning as two complementary frameworks," *International Journal of Tomography & Statistics*, vol. 6, pp. 122–127, 2007.

[162] ——, "Reinforcement learning versus model predictive control: A comparison on a power system problem," *Part B (Cybernetics) IEEE Transactions on Systems, Man, and Cybernetics*, vol. 39, no. 2, pp. 517–529, Apr. 2009.

[163] R. Negenborn, B. De Schutter, M. Wiering, and J. Hellendoorn, *Experience-based model predictive control using reinforcement learning.* TRAIL Research School, 2011.

APPENDIX A. ANALYSIS ON THE FAVORABLE FEATURE OF RPR AS THE VSM PREDICTOR

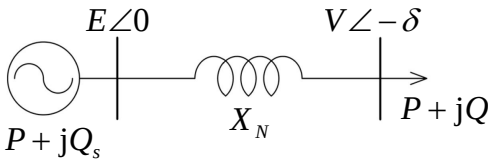


Figure A.1: Single machine - load 2-bus system.

For the 2-bus system shown in Figure A.1, under the voltage base E and the power bases E^2/X_N, the system equations in per unit quantity are [5]:

$$p = v \sin\delta \tag{A.1}$$

$$q = v\cos\delta - v^2 \tag{A.2}$$

$$q_s = 1 - v\cos\delta\,. \tag{A.3}$$

Given

$$q = p\tan\varphi$$

, where φ is the impedance angle of the load, for the high voltage solution branch we have

$$\left(p + \frac{1}{2}\tan\varphi\right)^2 + \left(q_s - \frac{1}{2}\right)^2 = \left(\frac{1}{2}\sec\varphi\right)^2, \tag{A.4}$$

or

$$q_s = 1/2 - \sqrt{1/4 - p\tan\varphi - p^2}\,, \tag{A.5}$$

where $1/4 - p\tan\varphi - p^2 = 0$ for the critical point of PV curve. Thus the system always goes to collapse when $q_s = 1/2$, no matter what the load it is. Further when $\varphi \geq 0$ is fixed, the p-q_s curve of the high voltage solution branch is just a section of quarter circular. Further notice that when

φ is fixed, VSM and RPR of this system are just the affine functions of p and q_s respectively. This implies that RPR is a good VSM predictor for the 2-bus system.

For multiple-machine system, according to [58], the system can be seen as approximately decoupled single-port Thévenin equivalent circuit from each load, when LID is fixed along the direction of the initial loads. The voltage of the jth equivalent source is $E_{eqj} = K_j V_G$, where K_j is a constant row vector which can be obtained from the bus-admittance matrix, and V_G is the column vector of voltages of the generators. Besides, under the assumption we can observe that the ratios among load bus voltages remain approximately constant with load increasing. In this case, it is trivial to show that the load current vector I_L can be written as

$$I_L = AI_G + BV_G\,, \tag{A.6}$$

where I_G is the vector of generator currents; A and B can be obtained from the bus-admittance matrix and the initial load currents. Consequently, we have

$$S_{eqj} = E_{eqj} I^*_{eqj} = E_{eqj} I^*_{Lj} = K_j V_G (A_j I_G + B_j V_G)^*\,, \tag{A.7}$$

where S_{eqj} is the complex generation of jth equivalent source. Let

$$\alpha = (V_{G1}/V_{G1}, V_{G2}/V_{G1}, \ldots, V_{Gn}/V_{G1})^\top\,, \tag{A.8}$$

where $\alpha_i = \angle\theta_{i1}$ is the voltage angle difference between the ith generator bus and the first generator bus; n_G is the number of generator buses. Then it is trivial to see

$$S_{eqj} = K_j \alpha \sum_{i=1}^{n_G} \frac{A^{\mathrm{H}}_{ji}}{\alpha_i} S_{Gi} + |V_{G1}|^2 K_j \alpha \alpha^{\mathrm{H}} B^{\mathrm{H}}_j\,, \tag{A.9}$$

where H indicates the conjugate transpose; S_{Gi} is the complex generation of ith generator bus. By some particular dispatch scheme, voltage angles of generators could be approximately constant; i.e. α is fixed. In this case, (A.9) implies that the RPR of each equivalent source is just an affine function of generators RPRs. Choose j to indicate the single port Thévenin equivalent circuit with lowest VSM. According to the implication of the 2-bus system, the RPR of the jth equivalent source, i.e., the affine function of generators RPRs, can be used as a VSM predictor. For real

system operation, E , X_N , LID, and α could vary within certain ranges. The conclusion here is not accurate but it arguably approximates the real situation.

APPENDIX B. PROCEDURES TO GET THE EMPIRICAL MEAN AND PREDICTION UPPER BOUND OF PREDICTION ERROR

This appendix devotes to get the blue and the black curve in Figure 3.6, which are used to estimate the prediction interval. Suppose we have obtained the prediction absolute error $|e_j|$ and the corresponding $locRMSE_j$ for all j during the validation, where j is the index for an operating point in the validation set. This finite set of samples cannot directly gives the empirical mean $\overline{|e|}$ or the empirical prediction upper bound PB at any given $locRMSE$, which is treated as a continuous variable in Figure 3.6. So we need to smooth the data over $locRMSE$.

The standardized tri-cubic kernel, similar to the one used in (3.9), can be used for smoothing. But instead of KNN window function, here we use fixed window width [90]. Therefore, the blue curve can be obtained by

$$\overline{|e|}(locRMSE) = \sum_{j=1}^{N^v} \omega_j(locRMSE) \cdot |e_j| \tag{B.1}$$

where $\omega_j\,(locRMSE)$ is the standardized kernel weight of point j with respect to $locRMSE$ as the neighborhood center (similar to z_0 in (3.9)). Using the same weights, standard kernel smoothing estimation can be applied to obtain the inverse cumulative distribution function of $|e|$ conditional to $locRMSE$, denoted by $F^{-1}_{locRMSE} : cl \mapsto |e|$, where cl is given prediction confidence level (95% is this work). In Matlab, this can be achieved by the routine *ksdensity*. Using these notations, we can formulate the black curve by

$$PB(locRMSE) = F^{-1}_{locRMSE}\,(0.95) \tag{B.2}$$

Choosing the best kernel or tuning the smoothing parameters is not a critical issue for our purpose, especially for this one-dimensional problem. So we omit such details.

APPENDIX C. FUNDAMENTAL OF MODEL PREDICTIVE CONTROL

Model predictive controllers originally rely on dynamic models of the target system, most often a linear ordinary difference equation (ODE):

$$x_{n+1} = A_n x_n + B_n u_n \tag{C.1}$$

$$y_n = C_n x_n \tag{C.2}$$

$$x_0 = X_0 \tag{C.3}$$

where $x_n \in \mathbb{R}^N$ is the state variable at time step n, X_0 is the known initial value of x_n, $u_n \in \mathbb{R}^M$ is the control input, and $y_n \in \mathbb{R}^L$ is the output at time step n. All these variables and the matrixes $A_n \in \mathbb{R}^{N\times N}, B_n \in \mathbb{R}^{N\times M}, C_n \in \mathbb{R}^{L\times N}$ could be time varying. A general optimal feedback control problem is to find a *control law* sequence $\pi = \{\pi_j\}_{j=0}^{D-1}$, where $\pi_j : \mathbb{R}^{L\times(n+1)} \mapsto \mathbb{R}^M$, such that the bounded feedback control $u_n = \pi_n((y_0, y_1, \ldots, y_n))$ minimize a convex cost function $f(x, u, D)$ within the control horizon $D \in \mathbb{N}^+$, where $x = (x_0, x_1, \ldots, x_D) \in \mathbb{R}^{N\times(D+1)}$, $u = (u_0, u_1, \ldots, u_{D-1}) \in \mathbb{R}^{M\times D}$ give the states and control inputs at all time steps within the horizon. f is usually of the quadratic form

$$f(x, u, D) = \sum_{n=0}^{D} x_n^\top Q_n x_n + \sum_{n=0}^{D-1} u_n^\top R_n u_n \tag{C.4}$$

where $Q_n \in \mathbb{R}^{N\times N}$ and $R_n \in \mathbb{R}^{M\times M}$ are positive semidefinite matrixes. Thus, the optimal feedback control problem for $n = 0, 1, \ldots, D$ can be formulated by

$$\begin{aligned}
\min_{\pi} \quad & \textstyle\sum_{n=0}^{D} x_n^\top Q_n x_n + \sum_{n=0}^{D-1} u_n^\top R_n u_n \\
\text{s.t.} \quad & \textit{for all } 0 \le n \le D-1: \\
& x_{n+1} = A_n x_n + B_n u_n \\
& y_n = C_n x_n \\
& x_0 = X_0 \\
& u_n = \pi_n\left((y_0, \ldots, y_n)\right) \\
& b_n \le u_n \le c_n
\end{aligned} \tag{C.5}$$

Comparing to open-loop control, engineers normally prefer closed-loop (feedback) scheme for general control problem since open-loop control doesn't take into account the unknown details of system and unforseen disturbances. The dynamics ODE included in the MPC scheme remedies this disadvantage to some extend by utilizing the knowledge of the system. However, the closed-loop solution still has the advantages that 1) sequentially updated output measurement makes the solution adaptive to unforseen disturbances, 2) once the control law is known (closed form solution or reduced to a simpler problem), we don't need to solve the original Problem (C.5) again for new initial states if the system structure is unchanged, and 3) the system is autonomous so that the stability of the controlling system can be analyzed. When closed-loop scheme is not available, we can simply consider the open-loop form of Problem (C.5):

$$\begin{aligned}
\min_{u} \quad & \textstyle\sum_{n=0}^{D} x_n^\top Q_n x_n + \sum_{n=0}^{D-1} u_n^\top R_n u_n \\
\text{s.t.} \quad & \textit{for all } 0 \le n \le D-1: \\
& x_{n+1} = A_n x_n + B_n u_n \\
& y_n = C_n x_n \\
& x_0 = X_0 \\
& b_n \le u_n \le c_n
\end{aligned} \tag{C.6}$$

A typical MPC applies the first several optimal control steps of Problem (C.5) or (C.6), $u_0^*, u_1^*, \ldots, u_P^*$ where $P < D-1$, in *receding horizon*: (a) at time step t, solving Problem (C.5)

or (C.6) for $n = t, t+1, \ldots, t+D$; (b) apply control $u_t^*, u_{t+1}^*, \ldots, u_{t+P}^*$ at time step $t, t+1, \ldots, t+P$ respectively; increase t to $t+P$ and replace X_0 by x_{t+P}, then go back to step (a). This procedure vividly describes the decision making process of human being in practice: we look ahead, make a plan, concentrate on executing the plan for some while, then look ahead again.