

Clustering Technique for Scenario Reduction in Post-Energy Transition Voltage Dips Assessment

R. Torkzadeh¹, J.B.M. van Waes², V. Čuk¹ and J. F. G. Cobben¹

¹ Department of Electrical Engineering, Eindhoven Technical University (TU/e),
5600 MB, Eindhoven, The Netherlands.

Phone number: +31 (0)40 247 6667, e-mail: [r.torkzadeh](mailto:r.torkzadeh@tue.nl), [v.cuk](mailto:v.cuk@tue.nl) and [j.f.g.cobben](mailto:j.f.g.cobben@tue.nl)@tue.nl

² TenneT TSO B.V. Utrechtseweg 310, Arnhem, The Netherlands.
e-mail: Jeroen.van.waes@tennet.eu

Abstract. The Dutch transmission system operator makes multiple scenarios to predict the future developments. These scenarios will help to define the risk factors and constraints in the grid, for which reinforcement planning is necessary. The developed grid after these reinforcements should continue to fulfil the power quality assessment criteria specified in the Dutch grid code. The reduction in system strength due to partial phase out of the conventional generation may have an adverse impact on the PQ, especially the voltage dips. Precise assessment criteria for voltage dips have been stipulated by the Dutch grid code that also need to be met after the energy transition. Evaluating all possible grid future scenarios can provide insight in possible future operating conditions. In practice, due to various combinations of network configurations, loading scenarios and dispatch scenarios, it is not possible to analyze all operating scenarios in detail. This paper presents a method to determine the most important scenarios for voltage dip assessments using a clustering technique. The proposed clustering technique reduces the number of scenarios that are needed to be assessed that makes the whole process doable in practice.

Key words. Energy Transition, Scenario Reduction, Power Quality, Voltage Dips, System Strength, K-means clustering.

1. Introduction

The energy transition will challenge the power systems which will be confronted by significant shifts, one of which is the shift from a dominant conventional synchronous generation (CSG) toward high penetration levels of inverter-based renewable energy sources (IBRs).

This shift will affect the system strength, particularly the short-circuit power at certain nodes of the system [1]. This is due to the limited capability of IBR in supporting short-circuit power versus CSG.

The second change driver that can alter the short-circuit power at various nodes is the grid reinforcements and reconfigurations to facilitate the energy transition. An example of reconfigurations is changing grid openings to control power flows. These could affect many aspects of the power system planning, operation, and control. One of the design aspects is power quality and in particular voltage dips.

The depth of the voltage dips and their propagation in future networks could significantly alter due to these changes. These parameters are determined by (local) short-circuit power which highly depends on generation dispatch.

Due to the generation shift, it is foreseen that voltage dips may become more severe and impact a higher number of clients.

It is important to build future scenarios to investigate this prediction. These scenarios should reflect possible future states of the grid for pre-defined planning horizons after the energy transition. In the Netherlands, TenneT TSO BV as the transmission system operator has developed an investment plan for 2020, which is referred to as IP2020 [2].

Various futuristic energy transition scenarios have been defined for a ten-year time horizon. Each energy transition scenario describes the energy network that would fulfil a specific energy transition target. The installed capacity of on-shore and off-shore windfarms, solar PV, number of EVs are some examples of details that are discussed within IP2020. Besides that, the IP2020 considers 2020, 2025 and 2030 as the short-, mid- and long-term horizons. Based on this document, three major scenarios can be distinguished:

- 1) climate agreement (CA)
- 2) alternative transition (AT)
- 3) fundamentals for system integration (FSI).

The CA contains the projects and grid development to fulfil the goals of the Paris Agreement. This evolution path is the most probable evolution of the grid. In addition to CA, two more evolutions have been considered. AT which focuses on the developments on the gas grid and the FSI which illustrates an extreme situation, with very high penetration of IBRs.

Table I shows the assumed installed capacity of various generation types for CA, AT and FSI for 2030 horizon. As it is illustrated, if FSI transition path happens by 2030, it is expected to have 57 GW of installed IBR. Considering the milestones and evolutions, six futuristic system states or “reference cases” could be considered as are depicted in Fig. 1 [3]. Each reference case contains information about the grid status for the corresponding year. For example, the CA-2030 reference case contains the details of the foreseen projects and developments including new connections, lines, substations, generation units, etc. for 2030.

Table I Installed capacity of generation per reference case [2].

Generation (GW):	CA	AT	FSI
Onshore wind	7.8	6	8.1
Offshore wind	11.3	10.6	14.6
Solar PV	25	14.3	34.3
Coal	0	0	0
Gas	8.9	8.9	7.5
Waste incineration and small CHPs	3.8	3.8	3.8
Hydrogen	0	0	1.4
Biomass	0	0	1.4

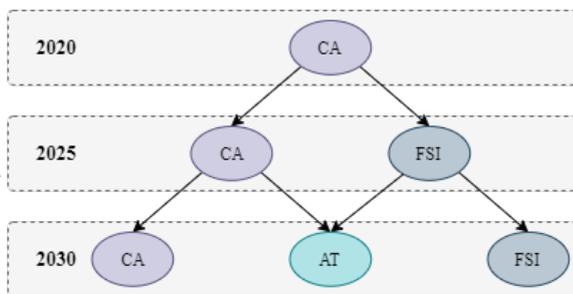


Fig. 1 Energy transition reference cases according to [2].

This information majorly illustrates the status of the infrastructure and the component rating (e.g. the installed capacity of CSGs), but not their operating points.

Per hour of each reference case a load and generation dispatch scenario is made based on market simulations. Market simulations calculate the dispatch of each type of generation and the power exchanges with the adjacent grids over HVDC and HVAC connections according to the forecasted load for each hour of a calendar year (52 weeks). This results in 8736 different snapshots per reference case per year.

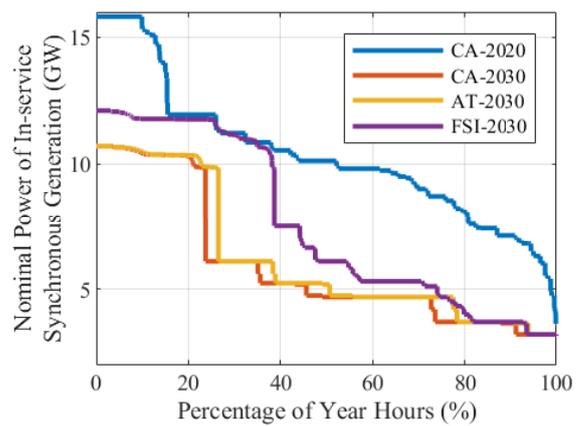
The renewable generation and load strongly depend on the weather conditions. The Netherlands logged weather condition for 1983, 2011 and 2012 are used, due to their characteristics. These three so called climate years (CYs) have been included for each horizon year.

Figure 2-a shows the duration curves of the nominal power of the in-service synchronous generation for CA-2020, CA-2030, AT-2030 and FSI-2030 (the CY2011 is assumed for all these curves). The load duration curves for aggregated load are depicted in Figure 2-b. Aggregated load is introduced in section 2-A. Figure 2-c illustrates the non-synchronous penetration index for each of the aforementioned scenarios. This index is introduced in section 2-A.

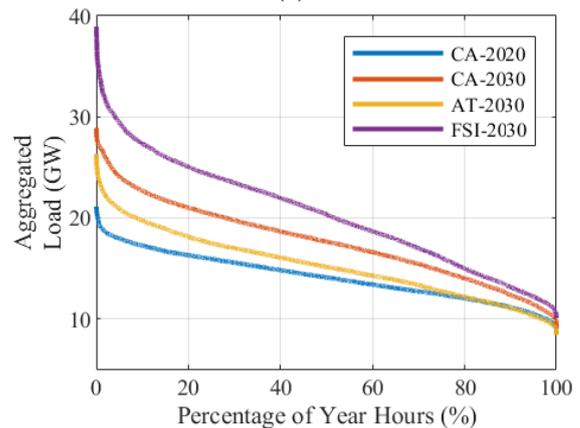
By comparing Figure 2-a and 2-b, it can be understood that however the loading is much lower in CA-2020 in comparison with all 2030 snapshots, still the amount of in-service synchronous generation for 2020 is much higher than the 2030 scenarios. Figure 2-c reveals the reason behind this difference. It's expected to have much more IBRs in-service in all the 2030 scenarios in comparison with CA-2020. For instance, non-synchronous penetration index of 60% and higher is realized only for 5 percentile of the year for CA-2020, however, for all the 2030 cases, we expect to have non-synchronous penetration index of 60% or above for more than 50 percentiles of a year.

Assessing all the hours of each reference case will help to cover many uncertainties in the planning studies, and risk assessment, but will burden massive computation and data processing as well.

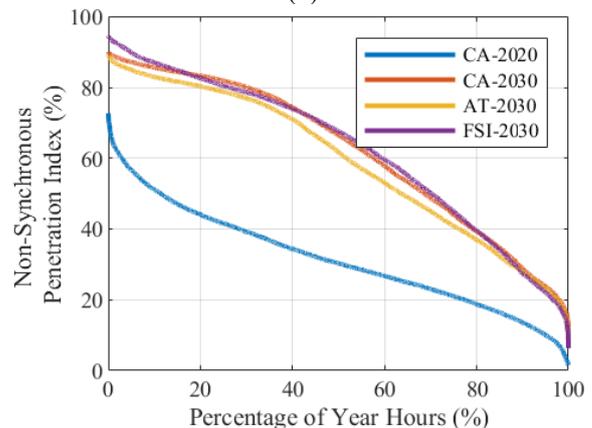
The aim is to identify the trend that is important for voltage dips. The origin of voltage dips are short circuits in the system such as failure of components, external causes, etc. These are in general not related to the operating conditions but occur randomly, so it is assumed that this input is the same for all scenarios. The depth and propagation of voltage dips are determined by the system strength. This is a local parameter. Locally there will always be deviations within a certain bandwidth. That is largely independent of the energy transition.



(a)



(b)



(c)

Fig. 2 The duration curves CA-2020, CA-2030, AT-2030 and FSI-2030 for a) nominal power of in-service synchronous generation, b) aggregated load and c) Non-synchronous penetration index.

The initial focus for voltage dips in the future is on the foreseen most *typical* states of the grid rather than *all* the plausible cases. Consequently, the question to be answered is how to reduce the number of snapshots that can represent all the hours of a reference case in a way to keep the computational burden as low as possible.

There is a limited number of studies that have been utilizing scenario reduction in power system problems [4]–[7]. The goal of scenario reduction in these works focused on the error of the wind forecasts and their impacts on the planning, however, none of them have presented a framework to minimize the number of hours/cases that are needed to be assessed for planning purposes.

Various techniques can be used for scenario reduction. An approach for scenario reduction is using data science techniques [8].

Amongst all these techniques, clustering methods are the ones, which can divide a large set of cases into a specific and limited number of subsets that their members share common features. In addition, clustering techniques unlike classification methods don't need labeling. It makes them more applicable in power system problems in which labeling is not always possible.

This paper addresses a clustering approach for scenario reduction which is suitable for voltage dips studies. The proposed clustering approach would lead to a limited set of candidates that represent a reference case.

The paper is organized as follows. Section 2 describes the methodology behind the work and discuss the K-mean clustering technique. The results are presented and discussed in section 3 and Section 4 concludes the paper.

2. Methodology

This paper considers the challenge of scenario reduction using clustering techniques. According to literature, two well-known clustering techniques can be used for this purpose [8]:

1. K-means algorithm, and,
2. Expectation-Maximization (EM) algorithm.

K-means is classified as a greedy algorithm, which tries to minimize the squared error of each point from its respective cluster *means* (centroid). K-means performs hard clustering, which means that each point in a d-dimension space is assigned to only one cluster.

EM is a generalized form of K-means. EM models the data as a mixture of normal distributions and attempts to find the cluster parameters (the *mean* and *covariance* matrix). This is obtained by maximizing the likelihood of the data.

The EM algorithm is considered as a soft clustering approach, in which, instead of making a hard assignment, it calculates the membership probability to each cluster for each point in a d-dimension space [8].

Considering the aim of this research, hard clustering is preferred. As the output of clustering should be a set of limited points that each one can represent all the members of the cluster that it belongs to. Hard clustering helps to highlight the differences between the clusters by defining certain boundaries for each cluster.

A. Indices for Clustering (Feature Definition)

The goal is to find typical operating conditions relevant for voltage dips by reducing the number of cases. Each reference case of Figure 1 contains 8736 operation conditions (hours) for each climate year. For the sake of simplicity, in this paper only one reference case and one CY will be assessed (i.e. CA2020-CY2011). To reach this goal, important parameters for voltage dips should be defined for each hour of each reference case.

These features are the ones that forms the n-dimensional space of the clusters. In this paper we proposed three indices to be used as these key features. These indices are defined in a way to distinguish between each hour with regards to the system state. In addition, as the most important parameters in voltage dips assessments are the ones that can impact the impedance of the network [9].

These indices are selected in way that represents the system state as well as the amount of in-service CSGs as one the

most prominent parameters that can impact the propagation of the voltage dips.

- 1) Aggregated load (in GW),
- 2) In-service synchronous generation (in GW), and
- 3) Non-synchronous penetration (in %).

The aggregated load is formed by summation of all the loads in the (E)HV network (P_{Load}^{HV}), the delivered power to MV grid (P_{Load}^{MV}), and the export over HVDC and HVAC to the adjacent grids (P_{HVDC}^{Export} and P_{HVAC}^{Export}) as shown in equation (1).

The in-service synchronous generation is calculated by summation of the nominal power of all CSG units that are in-service for that specific hour. Equation (2) shows how the proposed non-synchronous penetration index is calculated, this index majorly adopted from system non-synchronous penetration (SNSP) index introduced by [10]:

$$\text{Aggregated Load} = P_{Load}^{HV} + P_{Load}^{MV} + P_{HVDC}^{Export} + P_{HVAC}^{Export} \quad (1)$$

$$\text{Non-Sync. Penetration} = \frac{P_{RES} + P_{HVDC}^{Import}}{\text{Aggregated Load}} \times 100\% \quad (2)$$

Where P_{RES} is the summation of active power from RES generation units and P_{HVDC}^{Import} is the summation of imported power over HVDC.

B. The Proposed Scenario Reduction Algorithm

The overview of the proposed methodology for scenario reduction is presented in Figure 3. The first step is to build the hourly dispatch values for each of the generation units within the grid. The generation units contain all types of generation including CSGs and IBRs.

In order to define the per generation unit dispatch, a merit-order is needed. The merit-order defines which units have the priority to be dispatched first and defines the must-run units. The merit-order has a huge impact on the propagation of voltage dips. This is due to the fact that synchronous machines can supply short-circuit power and therefore the location of these machines is important in voltage dip assessment.

In addition, the minimum value for each generator to be in-service and maximum operational limits for active power are included in the merit-order. The output of this step will provide a $M \times N$ matrix, where $M=8736$ shows the hours within the year and N denotes the total number of generation units.

For each hour, the three feature indices are calculated and the $M \times N$ matrix is replaced with a 8736×3 matrix. The first column has the values for the aggregated load, the second column has the values for the total in-service CSGs, and the last column contains the values for the non-synchronous penetration index.

By applying the K-means clustering algorithm for different number of clusters from 1 to 100, the percentage of variance explained is calculated. In addition it essential to extract the most frequent dispatch patterns for CSGs, which is elaborated more in Section 2-D.

In this research, the minimum value for the number of clusters in which the 90% of variance is explained is chosen as the optimal number of clusters for that specific reference case (See Fig. 4-b).

By applying the most frequent dispatch patterns, the outliers of generation-mix are filtered out.

Finally, the closest hour to the centroid of each cluster is selected as the representative member of that specific cluster. The distance to centroid is calculated using the SSE explained by equation (3).

C. K-means clustering Algorithm

Given a dataset with n points in a d -dimensional space, and given the number of desired clusters k , the goal of representative-based clustering is to partition the dataset into k groups or clusters.

Given a clustering problem, a scoring function is needed to evaluate the appropriateness of this cluster. The sum of squared errors (SSE) scoring function is defined as equation (3).

$$SSE = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (3)$$

Where, k denotes the preferred number of clusters, x_j is the j^{th} point in the d -dimensional space, C_i is the i^{th} cluster and μ_i is the centroid of the cluster.

The goal of K-means is to find the clustering that minimizes the SSE score. More details on the K-means method can be found in [8].

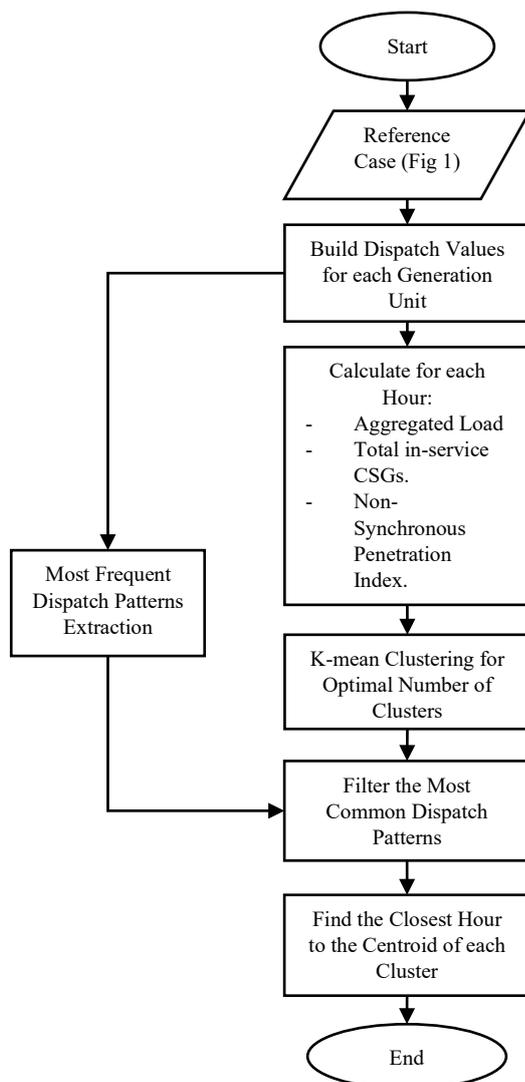


Fig. 3. The flowchart diagram of proposed scenario reduction technique.

D. Most Frequent Dispatch Patterns Extraction

As voltage dips are considered as local phenomena, system-level indices might not reflect the grid situation from the viewpoint of voltage dip assessment.

For a certain amount of total in-service CSGs, there might be various possible generation-mixes. The difference in these generation-mixes can impact the outcome of the voltage dips assessment. Therefore, it is essential to include an area-level index into the reduction algorithm. The procedure of most frequent dispatch patterns extraction is as follows:

1. All the non-zero arrays of $M \times N$ matrix are replaced with ones (1). The ones in a row indicate that for that specific hour which generation units are in-service. Zeros shows the unit was out-of-service.
2. All the columns related to non-CSG units are left out from the matrix. By considering the G as the total number of installed CSGs in the grid, $M \times N$ matrix is converted to a $M \times G$ matrix.
3. A string is made from the 1s and 0s in each row.
4. The number of repetitions of these strings in all rows show the frequency of each dispatch pattern.
5. In this work, the most frequent pattern(s) that is/are occurring for at least half of the year (cumulative sum of the repetitions should be greater than 4368) is/are selected as the most frequent dispatch patterns.
6. The hours related to extracted frequent pattern are defined to be used in the next step.

3. Results and Discussions

To facilitate the illustration of the proposed method, the output of the major steps discussed in Section 2-C are shown in Figure 4. Figure 4-a depicts all the operational points of CA2020-CY2011 for the proposed three indices (features). The hours with high penetration of RES and limited number of CSG are located on the upper right part of this figure.

As mentioned, to apply the K-means clustering technique, it is essential to define the optimal number of clusters. For this purpose the variance explained for 1 to 100 clusters is investigated. As it is illustrated in Fig. 4-b, the value for variance explained is more than 90% for 8 clusters. Consequently, all operating points of CA2020-CY2011 can be clustered into 8 groups. Figure 4-c shows them with different colors.

After applying the filter on most frequent dispatch patterns, various hours are filtered out as depicted in Fig. 4-d. Finally, Fig. 4-e shows the snapshots that are located to the closet vicinity of the centroid of each cluster. Table II shows the representative snapshots and their characteristics. It is important to mention that it is highly probable that each reference case has a different number of optimum clusters, therefore, comparing the reference cases becomes a matter of question.

In this work we suggest applying the voltage dips assessment to the reduced snapshots of each reference case. Then, compare the results of voltage dip assessments for each reference case. This comparison will show how each different evolution of the system can impact the voltage dips and their propagation through the system. Other methods like meshing the 3D space and comparing each section of the mesh for different reference cases is another alternative, which requires further investigation.

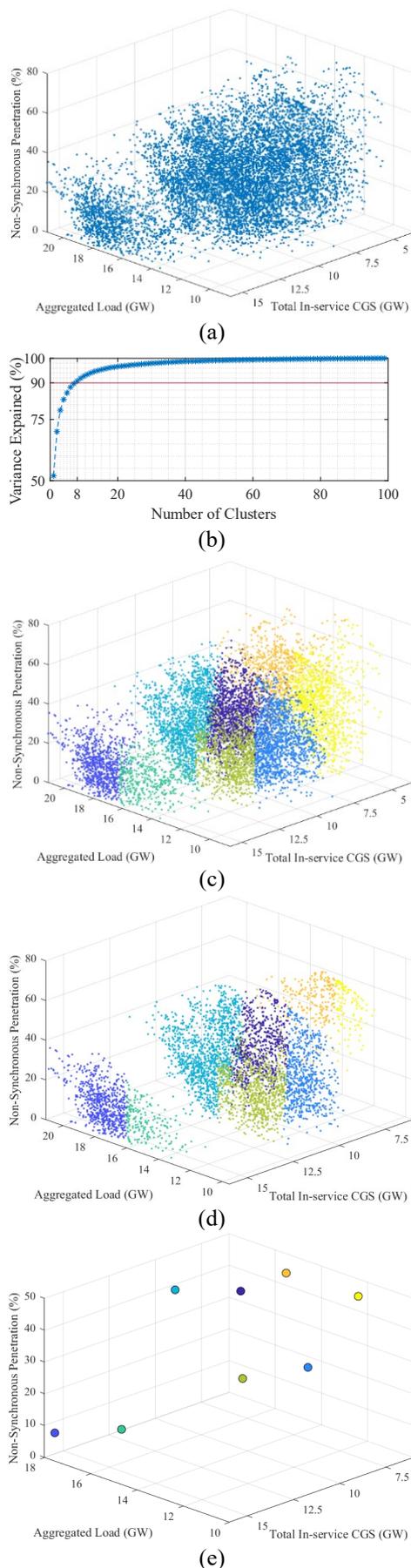


Fig. 4 Analysis for CA-2020-CY2011: (a) The scatter plot of the proposed indices for 8736 operation conditions; (b) elbow-curve to define optimal number of clusters; (c) the scatter plot for the clusters; (d) the clusters after applying the most frequent dispatch filter; (e) the reduced representative snapshots.

Table II The representative snapshots and their characteristics.

Snapshot Hour	Total In-Service CGS (GW)	Aggregated Load (GW)	Non-Sync. Penetration (%)
8719	9.802	14.461	48.7
394	15.824	17.703	8.1
7685	9.802	11.549	32.1
6300	10.845	16.470	46.1
1937	15.824	14.809	16.6
168	11.212	13.250	27.1
7969	7.122	14.636	48.4
1610	7.122	11.509	49.1

4. Conclusion

The foreseen changes in the networks due to energy transitions raise concerns that the power quality is deteriorating in the future. The reduced system strength could result in lower residual voltages and a larger area of propagation during voltage dips. Therefore, it is important to quantify these impacts. Various reference cases have been developed by the Dutch system operator that define future developments. These reference cases can be translated to 8736 snapshots for each horizon year.

The goal of this study is to reduce this enormous number of snapshots to a limited set of representative parameters, which is suitable for voltage dips assessments in future grids. This paper proposes a K-means clustering technique for this scenario reduction to find representative system states and minimize the computational burden and complexities. The results illustrate the effectiveness of the proposed method for the aim of scenario reduction.

Acknowledgement

This work is supported by Netbeheer Nederland. Many thanks to our TenneT colleagues for the useful technical discussions and contributions, especially to ir. Jorrit Bos.

References

- [1] B. Badrzadeh and Z. Emin, "System Strength," *Cigre Sci. Eng.*, no. 20, pp. 5–26, Apr. 2021.
- [2] TenneT and Gasunie, "IP2020 scenariobeschrijving," (in Dutch), 2019.
- [3] R. Torkzadeh, R. Peters, V. Cuk, J. van Waes, and S. Cobben, "Step-wise approach to investigate the impact of energy transition on voltage dips in the Dutch electricity grid," *Renew. Energy Power Qual. J.*, vol. 18, no. 18, 2020.
- [4] J. Li, F. Lan, and H. Wei, "A Scenario Optimal Reduction Method for Wind Power Time Series," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1657–1658, Mar. 2016.
- [5] J. Hu and H. Li, "A New Clustering Approach for Scenario Reduction in Multi-Stochastic Variable Programming," *IEEE Trans. Power Syst.*, vol. 34, no. 5, pp. 3813–3825, Sep. 2019.
- [6] Y. Feng and S. M. Ryan, "Scenario reduction for stochastic unit commitment with wind penetration," in *IEEE Power and Energy Society General Meeting*, Oct. 2014.
- [7] F. Luo, K. Meng, Z. Y. Dong, Y. Zheng, Y. Chen, and K. P. Wong, "Coordinated operational planning for wind farm with battery energy storage system," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 253–262, Jan. 2015.
- [8] M. J. Zaki and W. Meira, Jr, *Data Mining and Analysis: Fundamental Concepts and Algorithms*. 2014.
- [9] R. Torkzadeh, J. van Waes, V. Cuk, and S. Cobben, "Confidence Bound Assessment of the Dutch EHV / HV Grid Model Focusing on Voltage Dip Studies," in *14th IEEE PowerTech*, Jul. 2021.
- [10] Eirgrid, "System Non-Synchronous Penetration Definition and Formulation," 2018.