

Article

Power System Transition with Multiple Flexibility Resources: A Data-Driven Approach

Hao Li ^{1,2}, Ying Qiao ², Zongxiang Lu ^{2,*} and Baosen Zhang ^{3,*}¹ China Electric Power Planning & Engineering Institute, Beijing 100120, China; hli@eppei.com² Department of Electrical Engineering, Tsinghua University, Beijing 100084, China; qiaoying@tsinghua.edu.cn³ Department of Electrical and Computer Engineering, University of Washington, Seattle, WA 98195, USA

* Correspondence: luzongxiang98@tsinghua.edu.cn (Z.L.); zhangbao@uw.edu (B.Z.)

Abstract: Power systems are transitioning toward having high shares of variable renewable energy (VRE) with the help of flexibility resources. However, multiple flexibility resources on the generation, storage and demand sides introduce multiple technical and economic uncertainties, making the transition hard to predict. Moreover, the benefit of these resources in the transition is unclear. To fill these gaps, this paper proposes a data-driven approach to explore the transition to a high VRE share-oriented power system with multiple flexibility resources. This approach generates a wealth of possible transition paths under multiple uncertainties and then uses them to quantitatively analyze the transition. Specifically, the proposed method includes principal component analysis-based path visualization, multiple index-based transition milestone identification, cluster and distance calculation-based key influential factor identification, marginal index-based flexibility resource benefit comparison and Pareto frontier-based path recommendation. Case studies based on the Northwest China power system, which involves wind, photovoltaics and concentrated solar plants, validate the effectiveness of the proposed approach and further indicate that flexibility resources increase rapidly with the growth of the VRE share. Of the multiple flexibility resources, storage contributes the most. Key influential factors include the capital cost of VRE and storage along with coal price. These factors should be the focus in a low-cost and low-carbon transition.

Keywords: power system transition; flexibility resources; multiple uncertainties; data-driven; high renewable penetration



Citation: Li, H.; Qiao, Y.; Lu, Z.; Zhang, B. Power System Transition with Multiple Flexibility Resources: A Data-Driven Approach. *Sustainability* **2022**, *14*, 2656. <https://doi.org/10.3390/su14052656>

Academic Editor: J. C. Hernandez

Received: 19 December 2021

Accepted: 22 February 2022

Published: 24 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Power systems around the world are undergoing significant changes driven by the increasing availability of variable renewable energy (VRE). Many countries have proposed goals to achieve high levels of VRE penetration. The US is exploring scenarios in which renewable energy will supply 80% of demand by 2050 [1]. Europe is investigating the possibility of a 100% renewable power sector [2]. Similarly, China has indicated that its renewable generation share could reach 86% in 2050 [3]. However, these ambitious goals present challenges in power system planning and operations due to the variability and uncertainty of VRE [4,5]. The mismatch of VRE power supply and electricity demand needs flexibility resources to ensure an instantaneous power balance as well as long-term security of supply. Thus, ensuring that power systems have adequate flexibility resources is increasingly important [6–10]. In this paper, we explore the transition of power systems to high VRE penetration with the aid of flexibility resources.

With the increasing variability and uncertainty introduced by large amounts of VRE, simply using flexibility from the generation side is insufficient. “Combined flexibility—provided by all actors” becomes necessary [11,12]. However, multiple flexibility resources introduce uncertainties into system development. Flexibility resources come in many forms, including generation, demand and storage. Popular flexibility resources include thermal

unit retrofit [13], pumped hydropower, battery storage, compressed air energy storage (CAES) and demand response (DR) [7]. The maturity and resource potential of all of these technologies are uncertain, resulting in increasing uncertainty in power system transitions. Traditionally, these transitions are studied by presetting a small ensemble of development scenarios [14–17]. These scenarios are generally set by assuming different development levels of key influential factors, such as technology costs and technology efficiency. This method is suitable when there are only a few influential factors and the key factors are easy to identify. However, multiple factors containing uncertainty are introduced by various flexibility resources, and the most influential factors are unclear before an investigation. The traditional method with a few preset scenarios cannot consider all uncertainties, leading to an underestimation of the range of possible outcomes and further cognitive bias [18]. Therefore, coping with the increasing uncertainties from multiple flexibility resources is the key challenge when studying the power system transition towards a high VRE penetration target.

This uncertainty makes it challenging to reliably predict and analyze transition paths, and efforts have been made to address this challenge. Stochastic planning involves generating possible scenarios to represent uncertainty and conducting optimization for all scenarios [19–21]. However, this method suffers from the curse of dimensionality when coping with multiple uncertain factors related to multiple flexibility resources. Robust optimization can also consider uncertainty [22], but this method focuses on the most severe case and may result in conservative outcomes [23]. Recently, global sensitivity analysis (GSA) [24] and exploratory modeling [25], which randomly sample from uncertain factors and generate large numbers of possible outcomes, have been used to deal with multiple uncertain parameters. However, studies based on this idea used only simple statistical methods such as linear regression to analyze the outcomes, lacking the integration of popular data-driven techniques. These techniques, such as principal component analysis (PCA) and k-means clustering, have been proven to be effective when dealing with massive data [26], which can further facilitate transition path analysis for more insights. For example, existing studies present massive possible transition paths but lack milestone analysis [24,25]; thus, decision makers have difficulty directly obtaining insights from these paths. Additionally, the identification of key influential factors is currently based on single timescale input–output relationships [27–30]; time-varying traces are not considered. Additionally, although flexibility resources play an important role in power systems with a high VRE share, studies have addressed only some flexibility resources and calculated the benefits based on a few individual scenarios [17,31–34]. There is a lack of a quantitative benefit analysis and comparison that covers all generation-, storage- and demand-side flexibility and makes use of massive possible paths; hence, the research conclusions are not comprehensive.

In summary, existing studies fail to identify the transition pattern for the power system with multiple flexibility resources. The transition milestones, the key influential factors, the benefits of each flexibility resource, and the recommendations are not presented. To fill the gap, we propose a data-driven approach that can deal with multiple uncertain parameters and allow a comprehensive transition analysis considering multiple flexibility resources. The contributions of the proposed method are twofold:

- (1) We propose a data-driven approach that comprehensively analyzes the power system transition to a high VRE share. Compared with the existing studies, our approach can consider the uncertainties of multiple flexibility resources and explore the whole transition path space to obtain more reliable results. Additionally, the approach can present insights about transition milestones, key influential factors and the benefits of each flexibility resource; the recommendations are provided. The approach includes 5 steps: PCA-based path visualization, multiple index-based transition milestone identification, cluster and distance calculation-based key influential factor identification, marginal index-based multiple flexibility resource benefit comparison, and optimal path recommendation by Pareto frontier. In contrast to existing studies, the milestones include the gross and

marginal capacity requirements for flexibility resources under different VRE shares and the VRE share at which battery storage appears in the system and becomes the main flexibility resource. Cluster-based key influential factor identification can address time-varying transition traces rather than only a single timescale input–output relationship. The marginal benefits calculated from massive paths can facilitate the comparison of the technical, economic and environmental contributions of multiple flexibility resources.

(2) We provide a real-world study based on the Northwest China power system, which has the highest VRE penetration level in China. Studies demonstrate the effectiveness of the proposed data-driven approach for a high VRE share-oriented transition and further indicate that the requirement for flexibility resources is rapidly increasing. These resources have significant impacts on all technical, economic and environmental aspects, and the benefits from storage are most obvious. The most influential factors include the capital cost of VRE and storage, together with the coal price. These factors can facilitate a low-cost and low-carbon transition.

The remainder of the paper is organized as follows. Section 2 introduces the detailed methodology of the proposed data-driven approach. The results are analyzed in Section 3, with conclusions presented in Section 4.

2. Methodology

In this section, we first discuss the transition paths considered in this paper and then describe the proposed data-driven approach.

Power system transition paths can be described from multiple viewpoints, among which the most common are technology, the economy and the environment. The technical aspects considered in this paper are the mix of generation capacity, the mix of flexibility resource capacity and the VRE energy share (interchangeably referred to as the VRE penetration level in this paper). The economic aspect involves the system's total cost throughout the path and the cost breakdown. The environmental aspect refers to the annual carbon emissions and the total emissions throughout the path.

2.1. Path Generation under Multiple Uncertainties

The process of transition path generation is shown in Figure 1. The uncertainties are addressed by sampling from multiple uncertain parameters as inputs. Then, a multistage planning model considering generation and flexibility resources is built to optimize the transition path. Flexibility constraints such as thermal unit on/off status and DR regulation are included. All transition paths are set with the same VRE share target in the final year, which is similar to the target value set by governments [3,35]. Finally, transition paths from multiple viewpoints are generated.

2.1.1. Input

First, we determine the uncertain factors to be considered. Multiple technical and economic factors are regarded as uncertain inputs for transition paths, with an assumption that the policy environment is stable. Policy uncertainties can be addressed by setting scenarios such as those in reference [28]. Factors related to flexibility resources are included. Different from mature generation technologies, flexible resources are still in development; thus, the uncertainties of technical factors such as resource potential and technical efficiency are included.

Specifically, for generation technologies, the unit capital cost, fixed operation and maintenance (O&M) cost and fuel cost are assumed to be uncertain in the transition. The flexibility resources considered in this paper are thermal unit retrofit (i.e., reducing the minimum stable level), energy storage (i.e., pumped hydro, battery storage, and CAES, representing different storage capabilities, from a 2-h duration to more than a 16-h duration [36]) and DR (i.e., two types, load shedding and load shifting [37]). For flexibility resources, in addition to capital and fixed O&M costs, the compensation cost for consumers deploying DR is also considered. This is because compensation is an additional cost for

stakeholders such as generation companies and grid utilities. In contrast to that of generation resources, the potential of flexibility resources is hard to estimate since they are less coupled with natural resources. We address this issue by (1) setting three levels of thermal unit retrofit—no retrofit, medium retrofit and deep retrofit—as in Figure 2 (the maximum available reduction in units’ minimum stable level is assumed according to reference [9]); (2) assuming that the available investment capacity of battery storage and CAES linearly increases with time starting from the initial year but with uncertain growth ratios; and (3) setting the available DR investment capacity as an uncertain ratio of the peak demand. Moreover, the efficiency of different storage technologies is included as a technical parameter. Note that other flexibility factors, such as ramping rate and the minimum on/off time of units, are also considered in the planning model, but they are set to have no uncertainties because of the maturity of technology and studies showing they have less influence on flexibility [12,13].

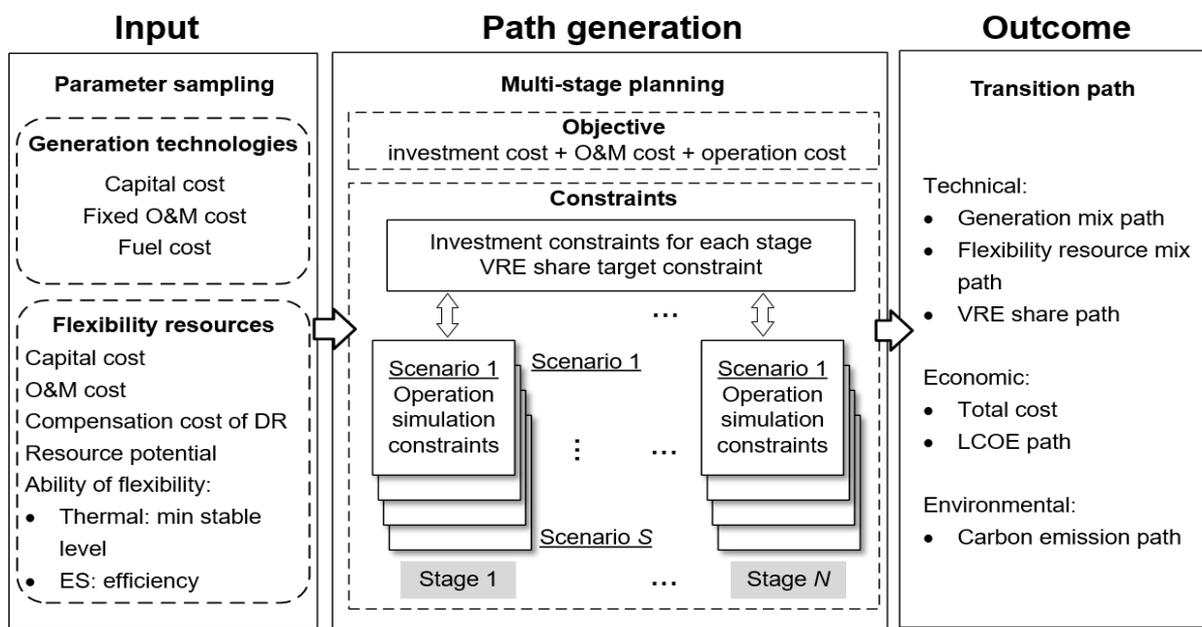


Figure 1. The process of transition path generation, where O&M is operation and maintenance, ES is energy storage, and LCOE is levelized cost of energy.

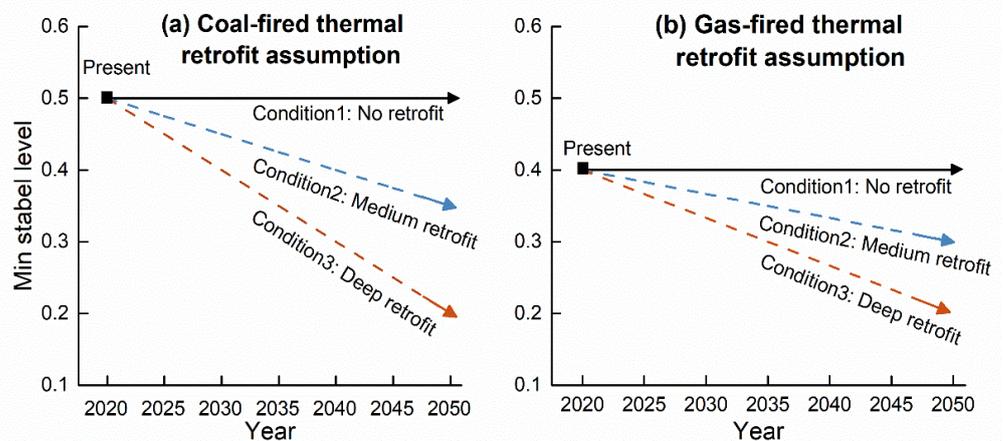


Figure 2. Different retrofit assumptions of thermal units’ minimum stable level. (a) Coal-fired thermal units; (b) gas-fired thermal units.

Then, the distribution of uncertain factors is needed for sampling. We assume that these factors are distributed uniformly, as is widely done in related studies [38]. The

comparison between different distribution assumptions will be conducted in future work. Finally, the parameters are randomly sampled and fed into the planning model.

2.1.2. Path Generation

In this paper, the whole transition period (2020~2050) is divided into 6 stages, with each stage representing 5 years. A multistage planning model is built to generate transition paths given the sampled inputs. The model incorporates both generation technologies and flexibility resources. The generation units include coal-fired and gas-fired thermal power, hydro power, wind power, PV power and concentrated solar power (CSP). The flexibility resources considered are thermal unit retrofit, pumped hydro storage, battery storage, CAES, load-shifting DR and load-shedding DR. The model is formulated as a bilevel stochastic optimization. Expansion decisions are made at the first level based on comprehensive economic optimization. On the second level, operation simulations are conducted for typical scenarios, and the estimated operation cost is returned to the first level. To optimize the whole transition path, we determine the expansion decisions of all stages simultaneously. For each stage, S typical scenarios (i.e., typical operation days) are considered for operation simulation. To ensure computational efficiency, the method in reference [39] is used, which aggregates the thermal units to address their on/off status without binary variables and formulates the planning model as a linear programming problem. The Northwest China system is denoted as five nodes corresponding to five provinces. Each node has unique characteristics in wind, PV, hydrology and demand, with the data (including typical scenarios) coming from reference [12]. The detail of the model can be found in Appendix A.

2.2. Data-Driven Path Analysis

With a large number of parameter samples fed into the planning model, various transition paths are generated. Then, the paths are analyzed in 5 steps: path visualization, milestone identification, key influential factor identification, flexibility resources benefit comparison and optimal path with a recommendation, as shown in Figure 3.

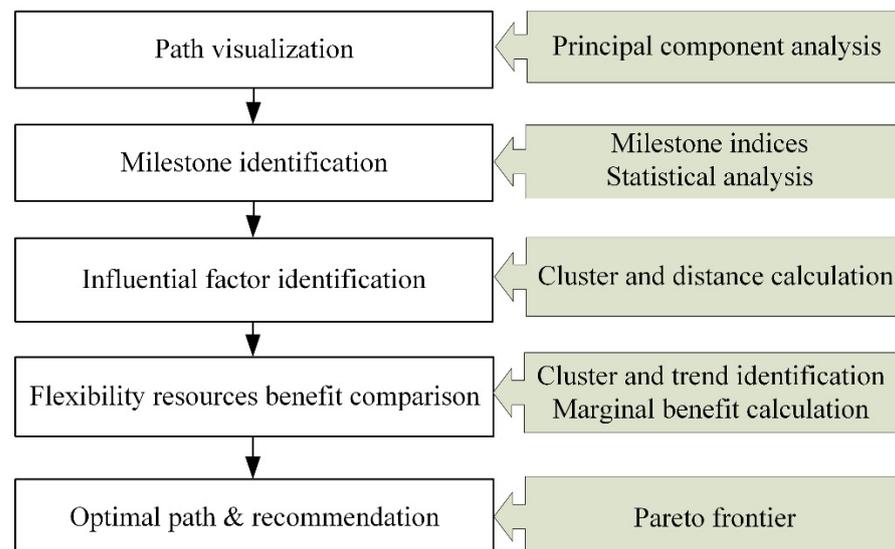


Figure 3. Proposed data-driven approach of the massive transition paths.

2.2.1. PCA-Based Path Visualization and Milestone Identification

The transition paths can be visualized by statistical plots of indicators such as generation capacity, VRE share, cost and carbon emissions. Notably, the generation mix is composed of many technologies and is thus highly dimensional, which requires dimension reduction methods for visualization. Dimension reduction can be performed using linear

methods such as PCA or nonlinear methods such as t-SNE. Compared with nonlinear methods, the linear method of PCA can identify the main components in high dimensions with a smaller computation burden. Thus, we use PCA [26] to linearly project high-dimensional points into low-dimensional space in this paper. Because a generation mix path consists of generation mixes at multiple stages, we first use PCA for each stage and examine the principal components that can represent the mix. Then, we aggregate the results of all stages to determine the representative components.

As the development of flexibility resources in the power system transition is not a focus of existing studies, the milestone indices proposed in this paper reflect this development. The gross flexibility resource requirements and the development of battery storage are both addressed. We focus on the battery because, as shown in many studies [7,40], battery storage is a promising flexibility resource and will account for a large portion of future resources. However, in the current stage, significant battery storage is not present in the Northwest China power system or in most systems around the world. Thus, it is worthwhile to single out the prospect of the battery in this system.

1. Milestones for flexibility resources

- Flexibility requirement (FR):

The flexibility requirement is the flexibility resource capacity \bar{I}^{FLEX} at a certain VRE penetration level ζ_0 :

$$FR = \bar{I}^{FLEX}(\zeta = \zeta_0) \quad (1)$$

where ζ represents the VRE penetration level.

- Marginal flexibility requirement (MFR)

The marginal flexibility requirement is defined as the flexibility resource capacity needed to improve the VRE generation share by 1 percent:

$$MFR = \frac{\Delta FR}{\Delta \zeta} \times 1\% \quad (2)$$

where ΔFR and $\Delta \zeta$ are the increase in the flexibility requirement and the increase in the VRE penetration level, respectively. The index varies with VRE penetration.

2. Milestones for battery storage

- Milestone of battery appearance (MoA):

The milestone of battery appearance is the VRE penetration level at which battery storage appears in the system:

$$MoA = \zeta_{MoA} \mid \begin{cases} \bar{I}^{BT}(\zeta < \zeta_{MoA}) = 0 \\ \bar{I}^{BT}(\zeta \geq \zeta_{MoA}) > 0 \end{cases} \quad (3)$$

where \bar{I}^{BT} represents the capacity of battery storage.

- Milestone of battery domination (MoD):

The milestone of battery domination is the VRE penetration level at which battery storage dominates in the system. Here, we define 'domination' as the capacity share of the battery among flexibility resources exceeding 50%. The index can be represented as

$$MoD = \zeta_{MoD} \mid \begin{cases} \bar{I}^{BT}(\zeta < \zeta_{MoD}) < 50\% \bar{I}^{FLEX} \\ \bar{I}^{BT}(\zeta \geq \zeta_{MoD}) \geq 50\% \bar{I}^{FLEX} \end{cases} \quad (4)$$

The collection of paths presents different values of milestones. The statistical results of the indices are presented in the results section.

2.2.2. Cluster and Distance Calculation-Based Influential Factor Identification

To identify key influential factors, several methods have been used in existing studies. Reference [41] changed some of the input and examined the local sensitivity. Reference [28] statistically found the subspaces within the uncertainty input space for different classes of paths. Reference [27] performed a multivariate linear regression and referred to standardized regression coefficients. Reference [29] calculated standard correlation coefficients to determine the correlations among the input and output. However, these methods mainly apply to inputs and outputs at a single timescale and cannot capture the relationship between time-series inputs and outputs.

To address this problem, we propose a distance-based method with the cluster technique to examine the relationship and identify the key influential inputs. The method is as follows:

- (1) Represent the i th input factor with \mathbf{y}_i . Since the factor value varies with development stages, \mathbf{y}_i is the time series composed of values at multiple stages:

$$\mathbf{y}_i = [y_{i,1}, y_{i,2}, \dots, y_{i,N}] \quad (5)$$

where $y_{i,1}, y_{i,2}, \dots, y_{i,N}$ represent the values of the i th input factor from stage 1 to stage N .

- (2) Cluster the collection of paths into several groups. Assume that the number of groups is R . The i th input factor of the paths belonging to the r th group corresponds to a subspace $\psi_i^{(r)}$ within its uncertain space Ψ_i .
- (3) For each input factor \mathbf{y}_i , calculate the mean value at each stage over subspace $\psi_i^{(r)}$ to obtain time-series mean values:

$$\bar{\mathbf{y}}_i^{(r)} = [\bar{y}_{i,1}^{(r)}, \bar{y}_{i,2}^{(r)}, \dots, \bar{y}_{i,N}^{(r)}], r = 1, 2, \dots, R \quad (6)$$

- (4) Standardize $\bar{\mathbf{y}}_i$ as

$$\tilde{\mathbf{y}}_i^{(r)} = [\tilde{y}_{i,1}^{(r)}, \tilde{y}_{i,2}^{(r)}, \dots, \tilde{y}_{i,N}^{(r)}], \text{ where } \tilde{y}_{i,n}^{(r)} = (\bar{y}_{i,n}^{(r)} - \mu_{i,n}) / \sigma_{i,n}, r = 1, 2, \dots, R \quad (7)$$

where $\mu_{i,n}$ and $\sigma_{i,n}$ are the mean value and standard deviation of the i th input factor at stage n over uncertain space Ψ_i , respectively.

- (5) Calculate the average distance among $\tilde{\mathbf{y}}_i^{(r)}$.

$$Dist_i = \frac{2}{R(R-1)} \sum_{r_1=1}^R \sum_{r_2=r_1+1}^R \|\tilde{\mathbf{y}}_i^{(r_1)} - \tilde{\mathbf{y}}_i^{(r_2)}\| \quad (8)$$

An input factor with a larger average distance means that the paths are more sensitive to this factor; hence, it is more influential than others. As illustrated in Figure 4, input factor A_1 has a larger average distance and thus is more influential than A_2 . In this way, we can compare and rank the input factors based on their average distance and identify key influential factors.

2.2.3. Cluster and Marginal Index-Based Flexibility Resources Benefit Comparison

In this section, we propose a method that can quantitatively compare the benefits of the transition of flexibility resources from the generation, storage and demand sides. The benefit is evaluated considering all of the technical, economic and environmental aspects of the transition. The general trend of the flexibility resources' impact on the transition paths is first identified, and the marginal benefit of these resources is then calculated.

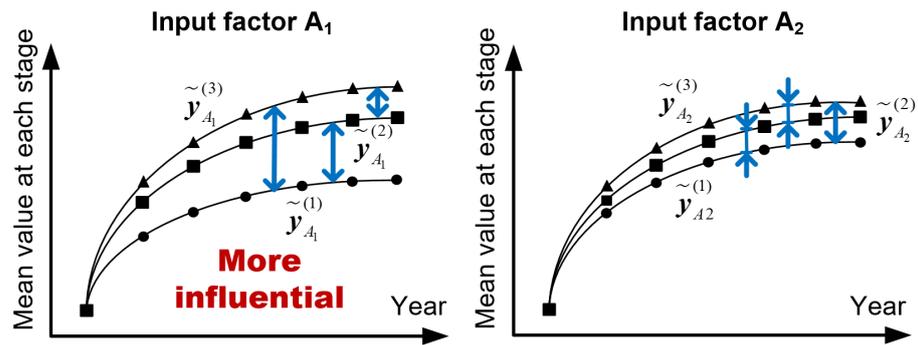


Figure 4. Illustration of distance-based key influential factor identification, where factor A1 has a larger average distance than factor A2.

1. Cluster and Trend Identification

The data representing transition paths from the technical aspect are composed of the capacities of different generation and flexibility technologies, which is high dimensional and makes it hard to observe the trend. Therefore, the cluster technique is used to represent the massive transition paths with typical paths so that the trend is identified by the analysis of typical paths. A technical path can be represented by a matrix as follows:

$$\underbrace{\begin{bmatrix} \bar{I}_{1,1} & \bar{I}_{1,2} & \dots & \bar{I}_{1,N} \\ \bar{I}_{2,1} & \bar{I}_{2,2} & \dots & \bar{I}_{2,N} \\ \dots & \dots & \dots & \dots \\ \bar{I}_{B,1} & \bar{I}_{B,2} & \dots & \bar{I}_{B,N} \end{bmatrix}}_{N \text{ stages}} \left. \vphantom{\begin{bmatrix} \bar{I}_{1,1} \\ \bar{I}_{2,1} \\ \dots \\ \bar{I}_{B,1} \end{bmatrix}} \right\} B \text{ technologies} \tag{9}$$

where \bar{I} represents the capacity of a certain generation or flexibility technology and B and N are the number of technologies and stages, respectively. The matrix can be transformed into a vector with the dimension $B \times N$, and the cluster can be conducted based on these vectors. We use the k-medoids method [26] for clustering rather than the most popular k-means method since k-medoids can directly obtain the typical path closest to the cluster center and avoid the distortion caused by extreme data.

The transition paths from the economic and environmental aspects can be described by a one-dimensional index (e.g., total cost, annual carbon emission); thus, the trend can be directly identified by scatter plots representing the relationship between the index and the flexibility resource capacity.

2. Marginal Benefit Calculation

To quantitatively compare the benefit of flexibility resources from different sides and make full use of the massive paths, marginal benefit indices are proposed in all technical, economic and environmental aspects.

With regard to the technical aspect, the marginal benefit of generation (MBG) is proposed, which is the capacity increment of a certain generation technology corresponding to the unit increase in flexibility resource capacity:

$$MBG = \frac{\Delta \bar{I}_{gen}}{\Delta \bar{I}_{flex}} \tag{10}$$

where \bar{I}_{gen} is the capacity of a certain generation technology gen , and \bar{I}_{flex} is the capacity of a certain flexibility resource $flex$.

With regard to the economic aspect, the marginal benefit of cost (MBC) is proposed to represent the impact on the transition cost of the unit increase in flexibility resource capacity:

$$MBC = \frac{\Delta C^{Total}}{\Delta \bar{I}_{flex}} \quad (11)$$

where C^{Total} is the total transition cost composed of three parts, as shown in Equation (1).

With regard to the environmental aspect, the marginal benefit of emissions (MBE) is proposed, representing the impact on annual carbon emissions by the unit increase in flexibility resource capacity:

$$MBE = \frac{\Delta E}{\Delta \bar{I}_{flex}} \quad (12)$$

where E is the annual carbon emissions.

To calculate the above indices based on the massive paths, linear fitting is conducted on scatter plots representing the relationship between the paths and the flexibility resource capacity. The fitted slope is the marginal benefit.

2.2.4. Pareto Frontier-Based Optimal Path Recommendation

To provide recommendations for stakeholders, the paths are evaluated from both economic and environmental perspectives. Paths with low costs and low carbon emissions are preferred. There is no obvious relationship between these two metrics. For example, a high coal price would reduce coal consumption and thus carbon emissions, but the change in variable operation costs could be in either direction. Therefore, to achieve low emissions and low cost simultaneously, the Pareto frontier of the paths is identified. Then, the collection of paths can be classified as optimal (on the Pareto frontier) or not. To reveal pivotal factors for optimal paths, a process similar to key influential factor identification is used. The policy for these factors can then form recommendations to stakeholders.

3. Case Study

This section presents the cost-optimal power system configurations of a high-VRE-penetration system from 2020 to 2050. The duration of one stage is 5 years. The VRE energy share target in 2050 is set as 80%, and 2500 paths for each thermal retrofit level are generated based on random sampling.

3.1. Input Data

We investigate the transition of the Northwest China power system. The system is composed of five provincial power systems, which have the highest VRE penetration level of all regional systems in China. It has long been dominated by coal-fired thermal power. In recent years, the wind and PV power capacities have increased continuously in Northwest China, and transforming the system into one with a higher VRE penetration level has been a great concern for the country. The demand and generation mixes for five provinces in the system in 2020 are shown in Figure 5, where wind and PV power account for 38.8% of the total generation capacity. The external load is the power transmitted to external power systems. In this study, we assume that the local load increases by 3% annually [40]. In Northwest China, the increase of external loads in the future will be covered by additional power plants, and these plants generally do not interact with the local grid. Since this study focuses on the local grid, the development of the external loads is not considered in the paper. The assumption of the determined load increase ratio can fix the demand for generation technologies and help us focus on the requirement of flexibility resources. The projection of the external load is another complicated issue that will be studied in the future.

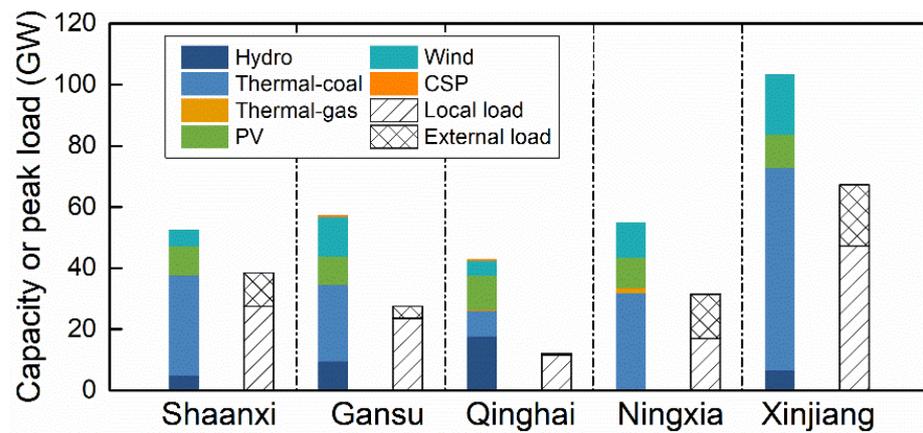


Figure 5. Peak demand and installed capacity for different generation technologies in the Northwest China system in 2020.

Based on the analysis in Section 2.1, we choose 21 uncertain input factors from technical and economic perspectives, as shown in Table A1. The range of factors at each stage is also shown. The range is determined based on references [12,39,40,42,43], which makes the projection considering the technology mature. The last column of Table A1 shows the reference of the range. When generating the paths, we sample all factors in all stages at each time and feed them into the planning model.

The available investment capacity for wind and PV are obtained from reference [39]. There is not much hydro power available in the system, so we set the hydro and pumped hydro potential based on reference [44]. The unit thermal retrofit cost is from reference [12].

3.2. Path Visualization

The transition paths under the deep thermal retrofit are shown in Figures 6 and 7. The paths under other retrofit levels are similar. In Figure 6a, each light blue line represents a VRE penetration path over time, and the dark blue line is the average trend of all paths. Figure 6b shows the distribution of the paths. Although all paths start at the same point and achieve the same VRE target in 2050, they vary over a large range under the considered deep uncertainty. The widest span appears in 2035, with the VRE share ranging from above 15% to below 60%.

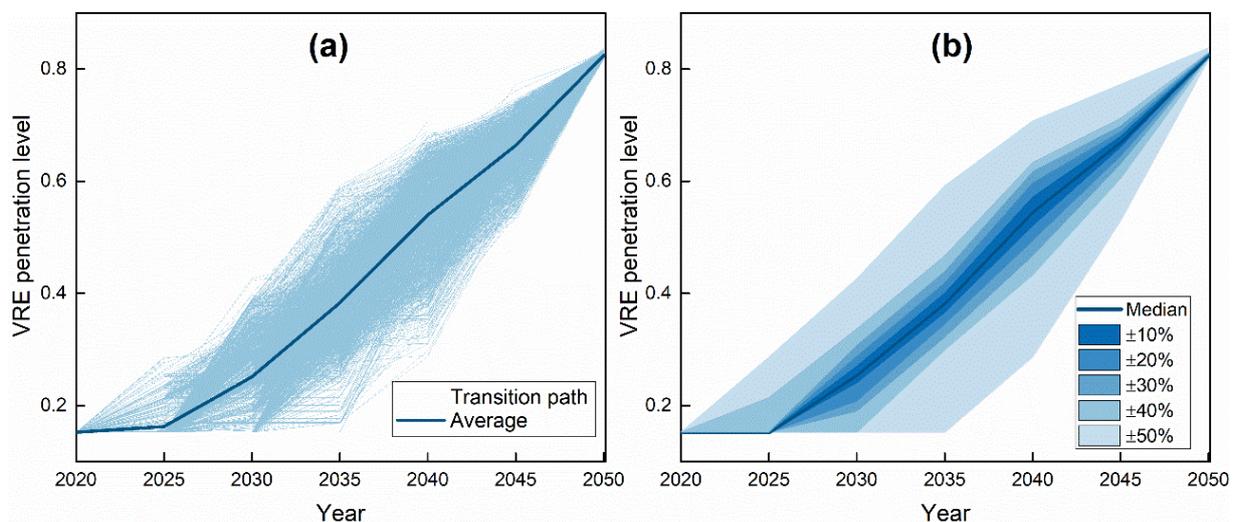


Figure 6. The transition paths of VRE penetration: (a) The VRE penetration paths; (b) The distribution of VRE penetration paths.

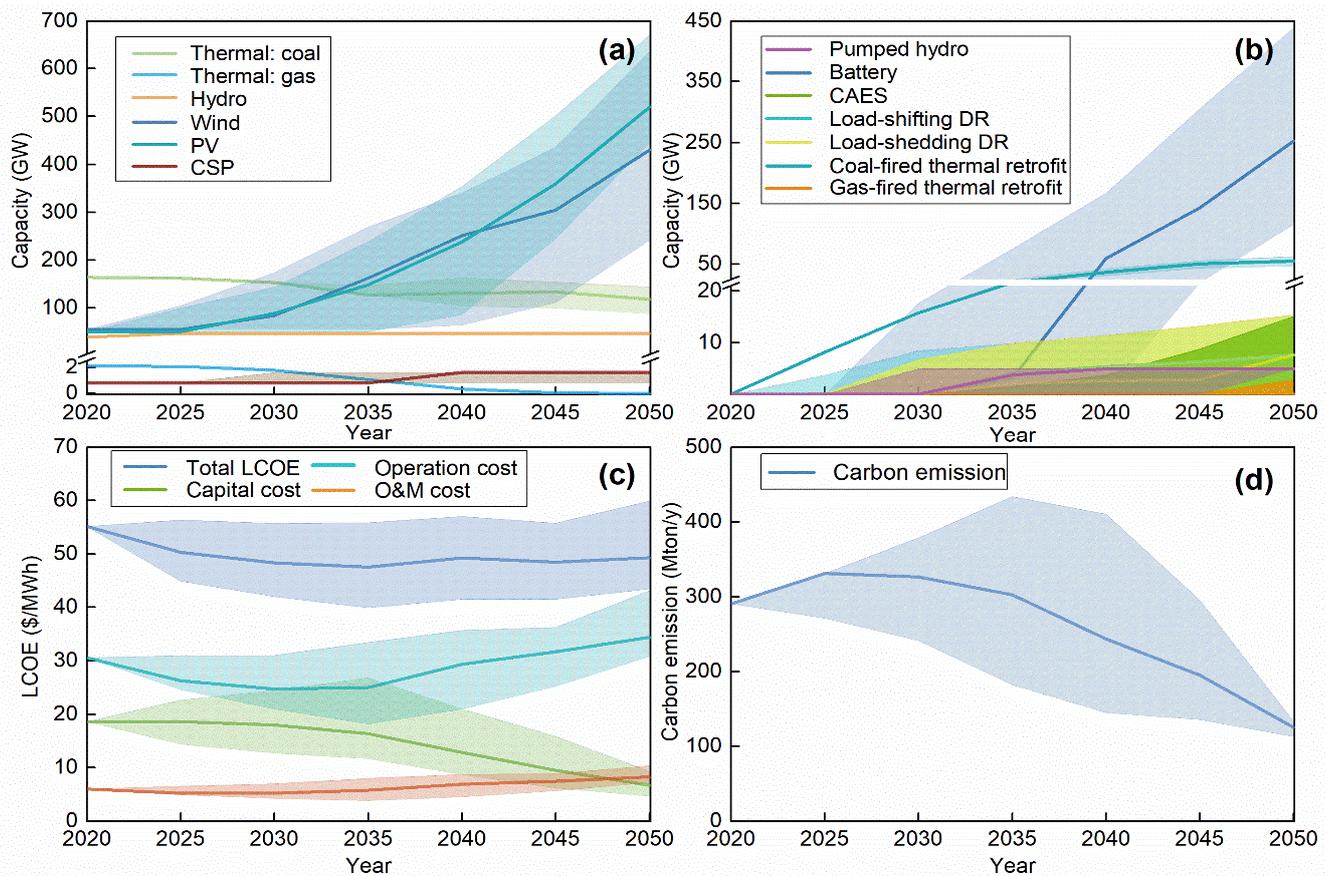


Figure 7. The transition paths and their distribution: (a) generation mix paths; (b) flexibility mix paths; (c) LCOE paths and the breakdown; and (d) carbon emission paths.

Figure 7 further shows the transition paths in the technical, economic and environmental aspects. The solid lines are the average values, and the shaded areas are the corresponding time spans. Figure 7a,b present the technical paths in terms of the generation mix and flexibility resource mix, respectively. Regarding the generation mix, wind and PV power increase continuously, and the average PV capacity exceeds wind capacity after 2040. Both coal-fired and gas-fired thermal units decrease. Hydro power remains stable, while CSP increases, accounting for a small portion. The results are relatively consistent with studies conducted by the government [3,35]. However, those studies failed to discuss flexibility resource allocation in detail. Among flexibility resources, battery storage dominates over a large span. Under the assumption of deep retrofit, coal-fired thermal retrofit ranks second in capacity, followed by the two types of DR. Due to the natural resource limit, pumped hydro is not largely allocated.

The levelized cost of energy (LCOE) is used to compare the cost and its breakdown in different years. It consists of capital cost, fixed O&M cost and variable operation cost:

$$LCOE = LCOE^{CAP} + LCOE^{OM} + LCOE^{OPR} \quad (13)$$

where $LCOE^{CAP}$, $LCOE^{OM}$ and $LCOE^{OPR}$ are the annualized capital cost, annual fixed O&M cost and annual variable operation cost divided by the annual generation energy, respectively. The capital cost sums up the products of capacity and annualized unit investment cost for all technologies in the system. The calculations of the fixed O&M cost and variable operation cost are presented in Section 2.1.2.

The system LCOE and its breakdown are presented in Figure 7c, where the capital cost increases but the variable operation cost drops. The total LCOE shows a slightly downward trend, which is also relatively consistent with reference [3]. As shown in Figure 7d, carbon

emissions slightly increase in the early stage and then decrease. This is because in the early stage, the emission reduction due to the slow VRE share growth cannot offset the emissions increase caused by the rising demand. According to the average path, the annual emissions will still increase by 12% and finally decrease by 57% under the 80% VRE energy share target.

We use PCA to visualize the high-dimensional generation and flexibility mixes. At each stage, the generation and flexibility mix is reduced to 2 dimensions, and the contribution of each technology to each dimension is shown in Table A2. It shows that two dimensions are enough to represent the high-dimensional generation and flexibility mixes, and the wind and PV contribute most to each dimension in all stages. Thus, we use the wind and PV capacity as two dimensions to illustrate the generation and flexibility mix along with the transition. Figure 8 shows the transition of the generation and flexibility mix denoted by wind and PV capacity. In 2050, there is a negative correlation between the two capacities because all samples in this stage have a similar VRE generation share, which means that the total VRE capacity is relatively determined. Additionally, the span of wind and PV capacity enlarges over time, indicating that the possible paths are increasingly diverse.

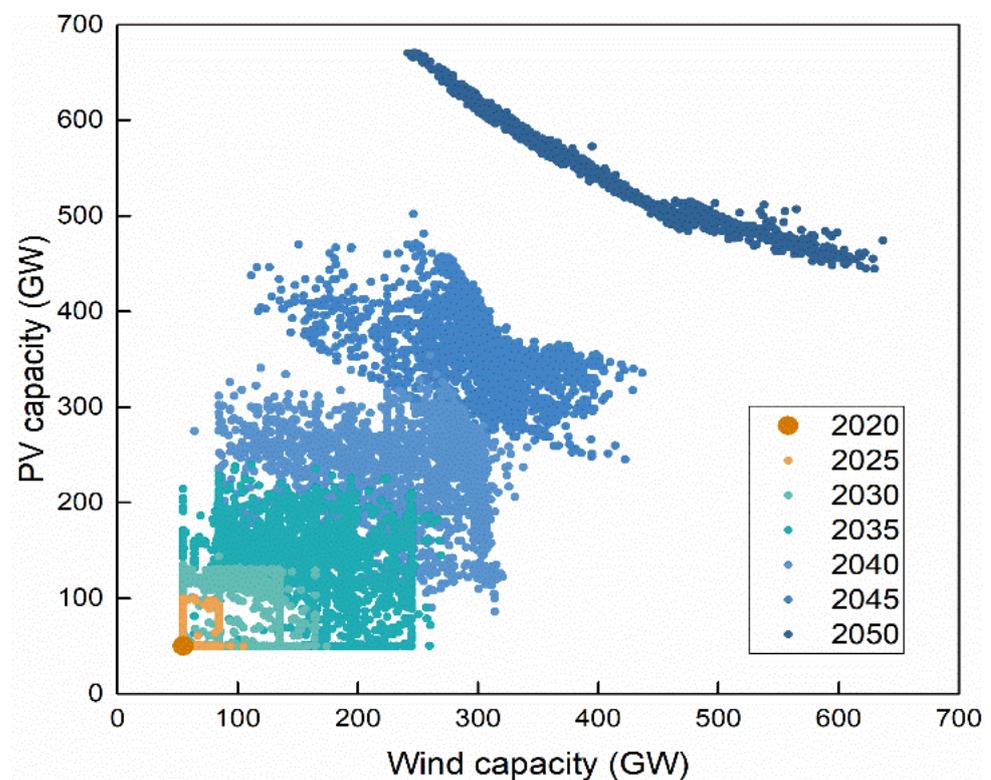


Figure 8. Visualization of the generation and flexibility mix transition.

We further compare the proposed data-driven approach with the traditional method that presets a small ensemble of scenarios to represent typical paths. We check its representativeness by comparing two paths. The first path is generated with the average input values among all samples, and the second is the average of the output paths from all samples. The results shown in Figure 9 reveal that the center of the input factor space does not correspond to the center of the output path space. Compared with the data-driven approach, the traditional method shows a maximum wind capacity deviation of 22.8%, a maximum PV capacity deviation of 41.1% and a maximum annual emission deviation of 9.3%. Therefore, preset scenarios with typical input values may not result in typical outputs.

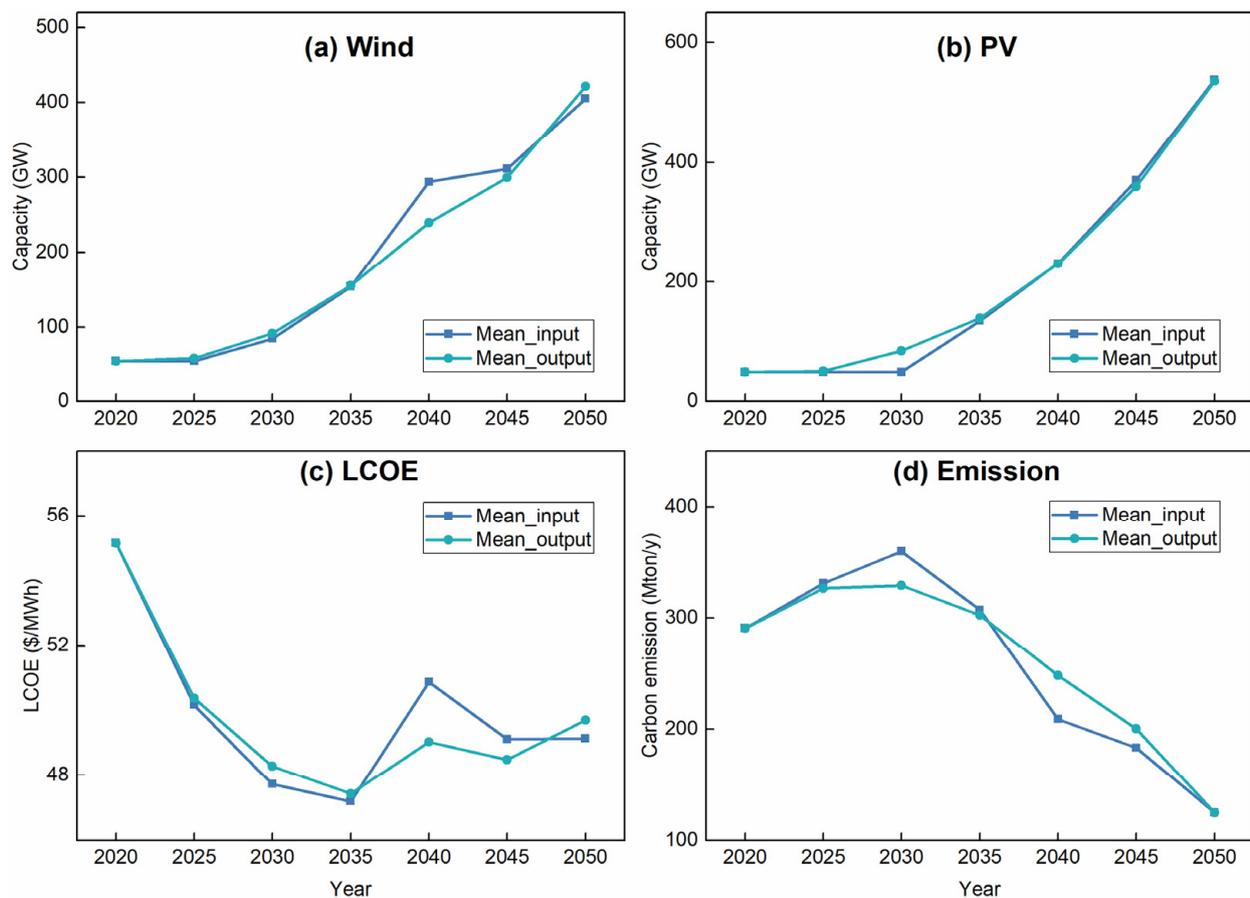


Figure 9. The comparison of the path with the average input values (labeled Mean_input) and the average path among all samples (labeled Mean_output) in terms of (a) wind capacity, (b) PV capacity, (c) LCOE, and (d) carbon emission.

The above results demonstrate that multiple uncertainties in the transition lead to diverse transition paths in the technical, economic and environmental aspects and thus cannot be ignored. Traditional methods preset a few possible scenarios without knowing the actual path space under multiple uncertainties. In contrast, the proposed data-driven approach can explore the path space to obtain more reliable results.

3.3. Transition Milestones

Figure 10 shows the flexibility requirement with respect to the VRE penetration level, together with the marginal flexibility requirement. The paths are under the deep retrofit. With the rise of the VRE penetration level, not only does the flexibility capacity increase, but more marginal flexibility resources are also needed. The flexibility capacity when the VRE share is 70~80% is approximately 23 times the capacity when the share is 10~20%. For the marginal flexibility requirement, the increase is approximately 11 times. This indicates that it is harder to further improve the penetration when it is at a high level, and the role of the flexibility resource becomes more important as the VRE share increases.

Regarding the battery requirement, Figure 11 shows the VRE penetration level at which the battery appears in the system or dominates the flexibility resources. The battery is most likely to appear in the system when the penetration is 40~45% and to dominate the flexibility resources at the penetration level of 65~70%.

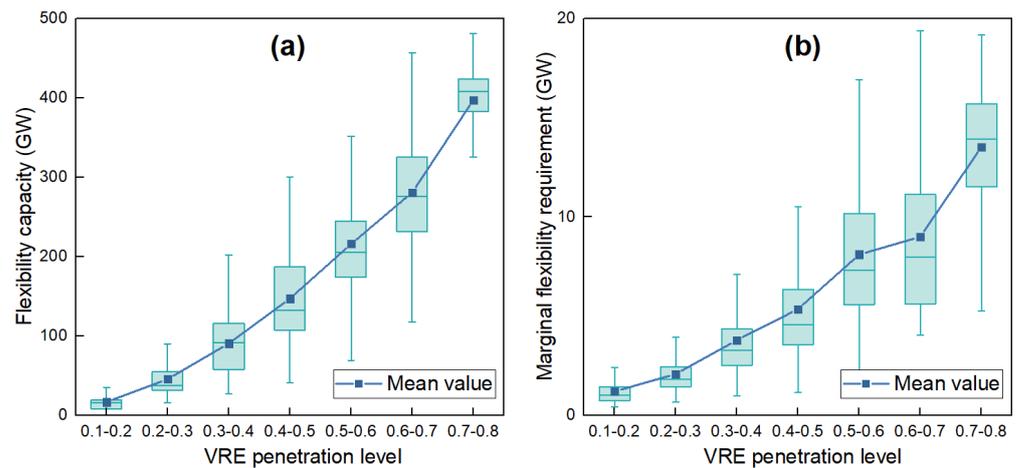


Figure 10. Flexibility requirements with respect to the VRE penetration level. (a) Flexibility capacity; (b) marginal flexibility requirement.

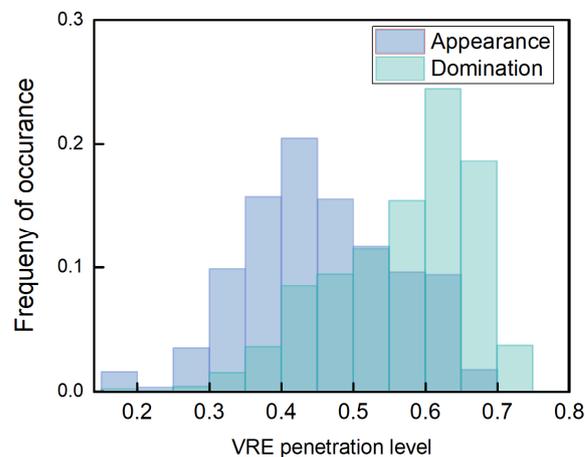


Figure 11. The VRE penetration level at which battery storage appears and dominates.

3.4. Key Influential Factors

The transition is characterized by multiple aspects, and the key influential factors for each aspect may be different. First, we explore the influential factors for the paths of VRE penetration to demonstrate the process of identification. Then, the influential factors for other aspects are identified.

For VRE penetration paths, we examine those under deep retrofit as an example. We cluster the paths into 4 groups. The lines in Figure 12 represent the average paths of each cluster. The figure reveals the four patterns of the transition paths in terms of VRE penetration, with a low penetration level (Cluster 1) or a high penetration level (Cluster 4) in most stages, medium penetration level with a high increase rate in the middle stage (Cluster 2) and a medium penetration level with a low increase rate in the middle stage (Cluster 3).

Within each cluster, we calculate the time-series means of each input parameter and normalize them, with the results shown in Figure 13. Some parameters, such as the unit capital cost of VRE and coal price, show larger differences across clusters, which means that the VRE penetration paths are more sensitive to these parameters. The average distance for each parameter is calculated and ranked, as shown in Figure 14. The most influential factors for the VRE penetration paths are the coal price, unit capital cost of wind power, available investment of the battery at each stage, unit capital cost of PV power and battery, which are consistent with the results in Figure 13.

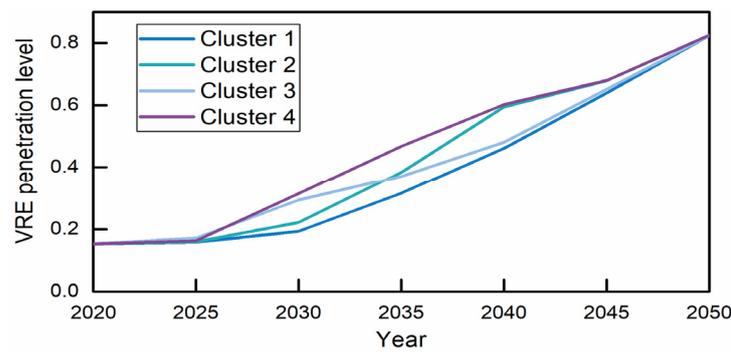


Figure 12. The clustered VRE penetration paths.

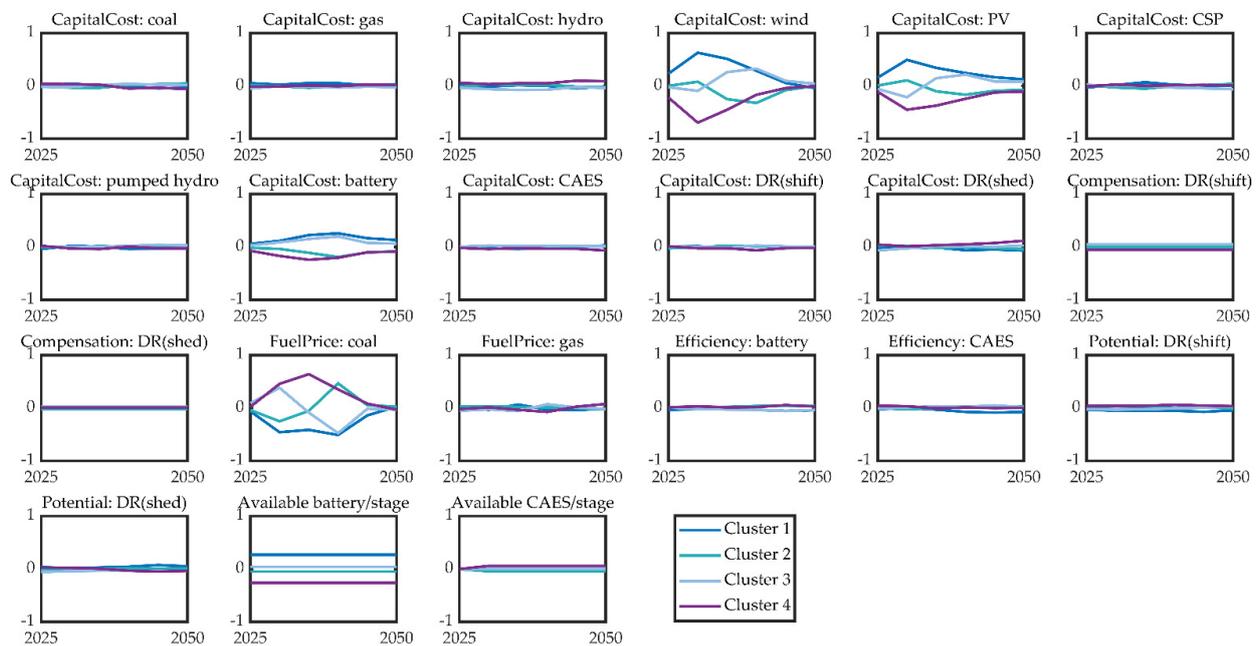


Figure 13. Normalized mean values of input parameters for different clusters of VRE share paths.

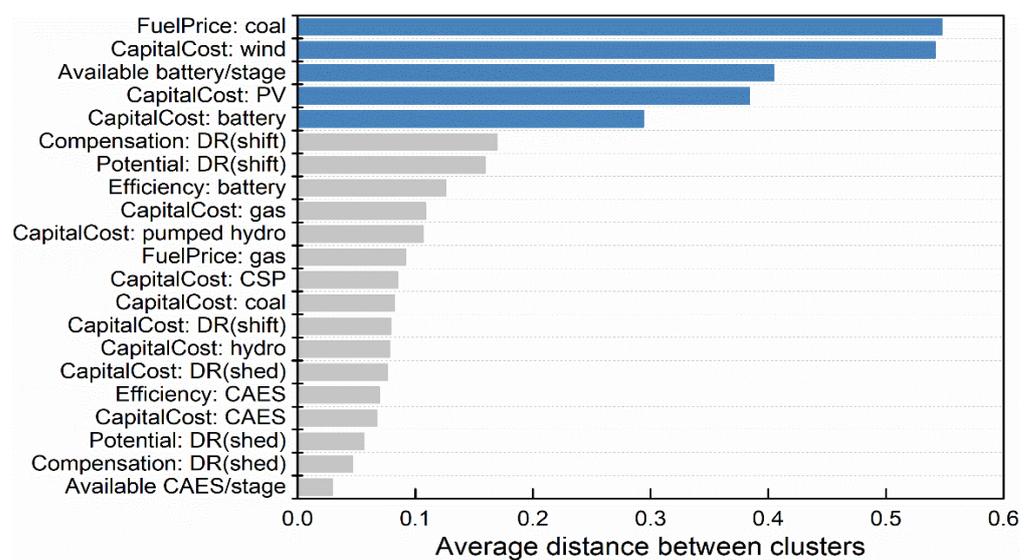


Figure 14. The average distance of input parameters between clusters of VRE generation paths.

Using the aforementioned process, we further examine the key influential factors for other aspects. The average distances of the input parameters are shown in Table A3. For all aspects, the key influential factors are similar, including the unit capital cost of VRE and battery, the coal price and the available battery investment capacity at each stage. However, the ranks of the parameters differ slightly. With regard to the technical aspect, the battery's available investment capacity at each stage is the most dominant, which indicates that the battery is a cost-effective choice and that its allocation is closely dependent on the available investment capacity. With regard to the economic and environmental aspects, the most important factor is the coal price, while other factors, such as the battery's available investment capacity, have different effects.

3.5. Benefit of Flexibility Resources

3.5.1. Technical Benefit

In this section, we focus on the impact of flexibility resources on the generation mix by considering clustered scenarios and statistical results. First, we cluster the paths of the generation mix into 2 groups and examine the typical scenarios of generation mixes and investment capacity mixes (Figure 15). Comparing these two scenarios reveals that the relative ratios among wind capacity, PV capacity and battery capacity are different.

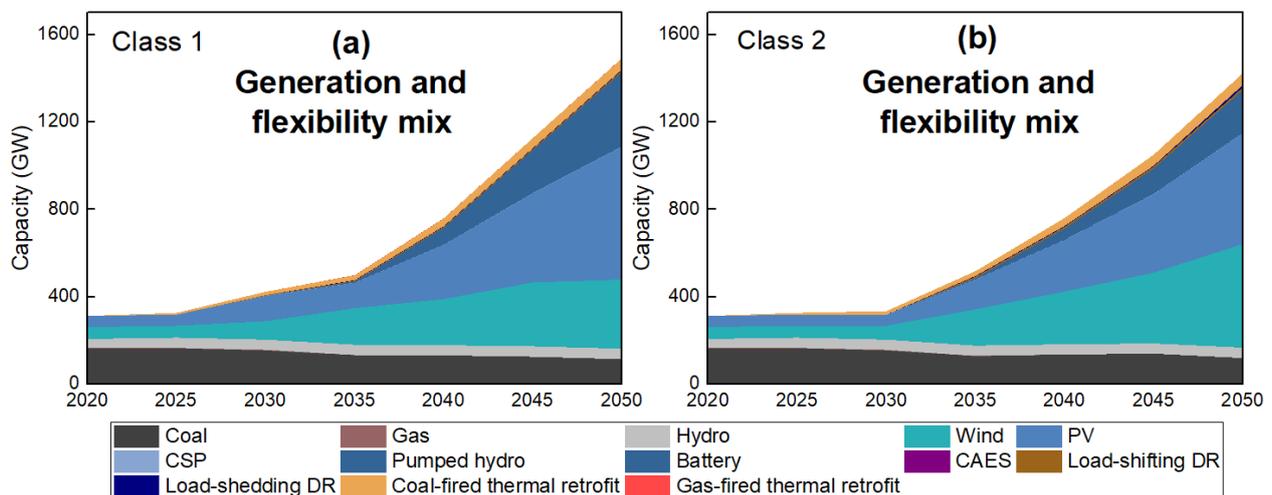


Figure 15. The generation mixes and investment capacity mixes of two clusters. (a) The generation mix for the first cluster; (b) the generation mix for the second cluster.

To further examine the relationship between the battery and generation mix, we analyze the statistical results for all samples (Figure 16). The capacity of battery and PV shows a linear positive correlation in most stages because storage is suitable for PV generation, which has a typical daily pattern that peaks at noon and drops to zero after sunset. However, the correlation is not obvious between battery and wind capacity. The association between battery and wind capacity in 2050 is mainly due to the linearity between PV and wind in this stage, as shown in Figure 8. The slopes of the linear fitting in Figure 16a reveal that an increase of 1 MW in battery capacity corresponds to approximately 1 MW in additional PV capacity.

3.5.2. Economic Benefit

We explore the economic impacts of the flexibility resources from the generation, storage and demand sides, as shown in Figure 17.

- (1) From the generation side, the total cost, including the investment, fixed O&M and variable operation costs for the whole transition, is compared across different thermal unit retrofit levels. The deeper the retrofit level is, the lower the total cost will be. We also calculate the marginal cost reduction as the cost reduction per retrofit capacity.

- For the medium and deep retrofits, the average marginal cost reductions over paths are 216 k\$/MW and 234 k\$/MW, respectively, which are close to each other.
- (2) From the storage side, the total cost shows a negative correlation with the storage capacity in 2050, which means that storage can also reduce the total cost. The slope of the linear fitting indicates that the average marginal cost reduction of the storage is 216 k\$/MW, similar to that of the thermal retrofit.
 - (3) From the demand side, there is no obvious correlation between the DR capacity in 2050 and the total cost, indicating that DR contributes little to the cost reduction. This is because, as shown in the table of the input parameters (Table A1), DR's unit compensation cost obtained from a study in China [45] is overall higher than the coal price. Therefore, the cost of deploying DR may be more than the saved fuel cost, making it less economical. Note that the DR results apply only to the considered system and data. For other countries where DR is more economical, the results may be different.

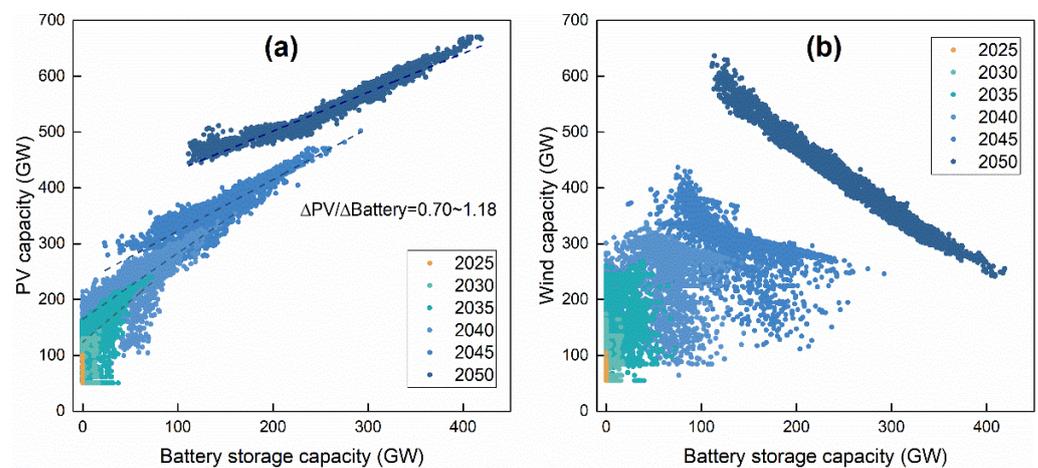


Figure 16. (a) The relation between battery storage capacity and PV capacity; (b) the relation between battery storage capacity and wind capacity.

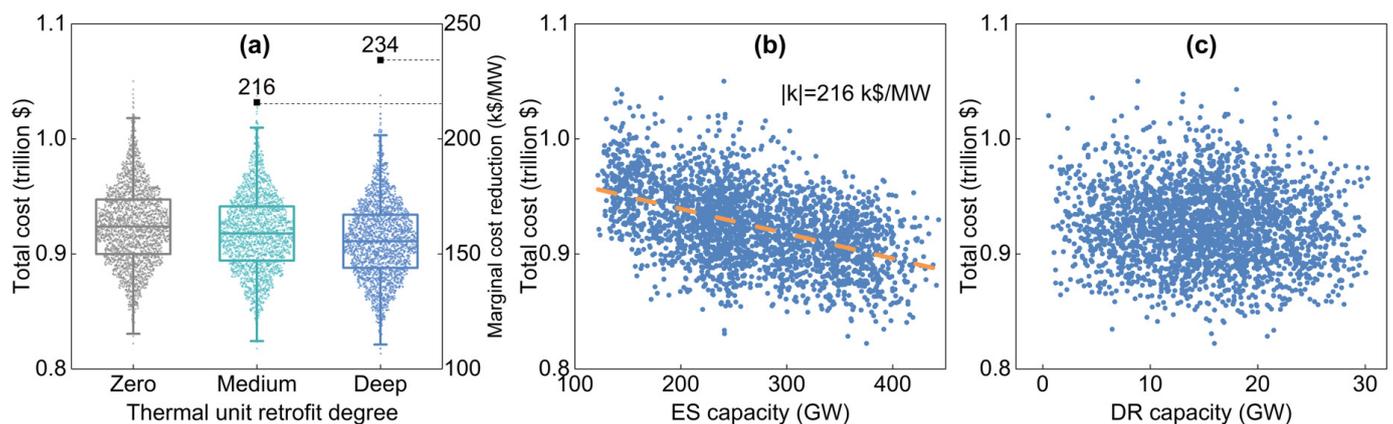


Figure 17. Impacts of flexibility resources on the economic transition paths. (a) The total cost and the marginal cost reduction of different thermal retrofit levels; (b) the total cost with respect to ES capacity; (c) the total cost with respect to DR capacity.

3.5.3. Environmental Benefit

In this section, we also explore the environmental impacts from the generation, storage and demand sides, as shown in Figure 18.

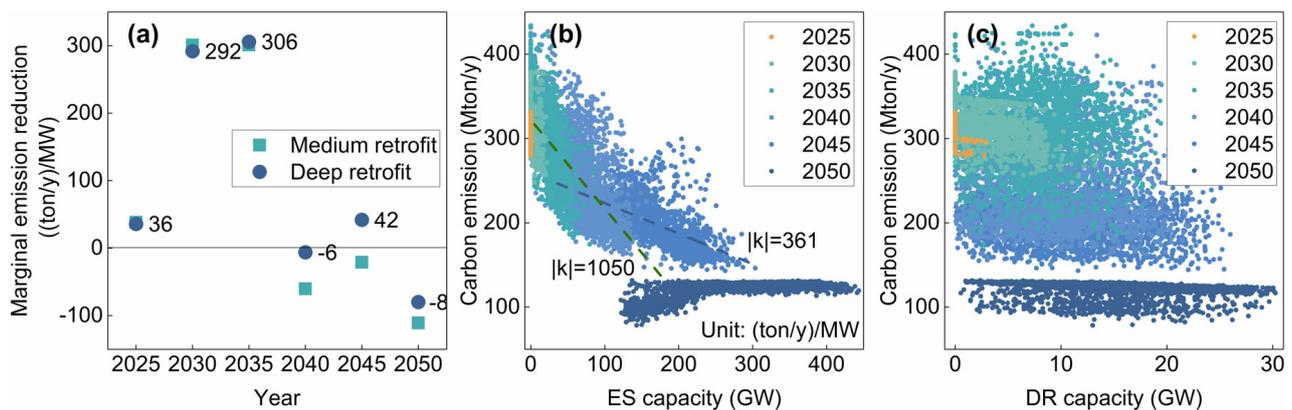


Figure 18. Impacts of flexibility resources on environmental transition paths. (a) The marginal emission reduction of the thermal retrofit; (b) annual emissions with respect to ES capacity; (c) annual emissions with respect to DR capacity.

- (1) From the generation side, the marginal emission reduction of the retrofit capacity is large in the early stages but negative in some stages. We investigate the reason by randomly checking a sample, as shown in Table A4. The table shows that the retrofit, which is cheaper than the battery in terms of investment cost, can replace part of the battery capacity under deep retrofit. This saves on the investment cost but may increase the operation cost because the thermal units are more likely to be committed to providing flexibility, which may result in more carbon emissions compared to the case with no retrofit but a large battery capacity.
- (2) From the storage side, the carbon emissions show a negative correlation with the storage capacity at most stages. In the last stage of 2050, emissions are relatively stable with different storage capacities because the VRE generation share in this stage is almost the same for all paths. From the slopes of the linear fitting, the marginal emission reduction of the storage is 361 (ton/y)/MW or above, higher than that of the thermal retrofit.
- (3) From the demand side, as discussed in Section 3.5.2, the allocation of DR may be influenced by complex factors, and thus, it does not show a clear relationship with the total cost or carbon emissions. We want to note again that the result may be different for other systems and data.

3.6. Optimal Paths and Recommendations

We evaluate the paths by their total cost and carbon emissions throughout the transition. First, all paths are plotted with respect to these two dimensions. Figure 19 shows the lack of a linear correlation between cost and emissions because complex factors influence them in different ways, as discussed in Section 2.2.4. Then, we identify the Pareto frontier of the paths as the optimal paths, which are shown in light blue in Figure 19. Finally, with a similar process for identifying key influential factors, we obtain the normalized mean values of input parameters from the optimal paths and compare them with those from nonoptimal paths. Figure 20 shows that the optimal paths generally require a lower unit capital cost of wind, PV and battery, a lower coal price and a higher available battery investment capacity at each stage. Stakeholders should focus on these factors. Under an early stage, the coal price should be focused on since it composes a large part of the system LCOE when the VRE share is not high. Then, policies should be proposed to encourage the development of storage because its lower cost and high capacity contributes to the optimal transition paths.

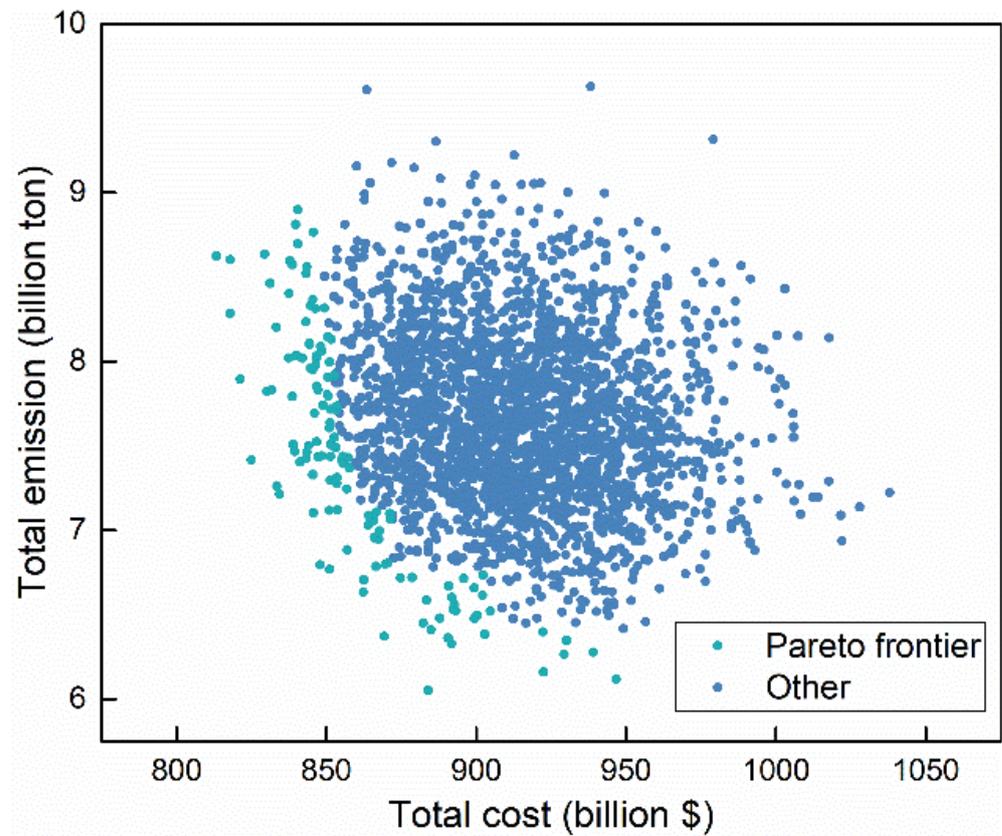


Figure 19. The samples and the Pareto frontier in terms of total cost and emission.

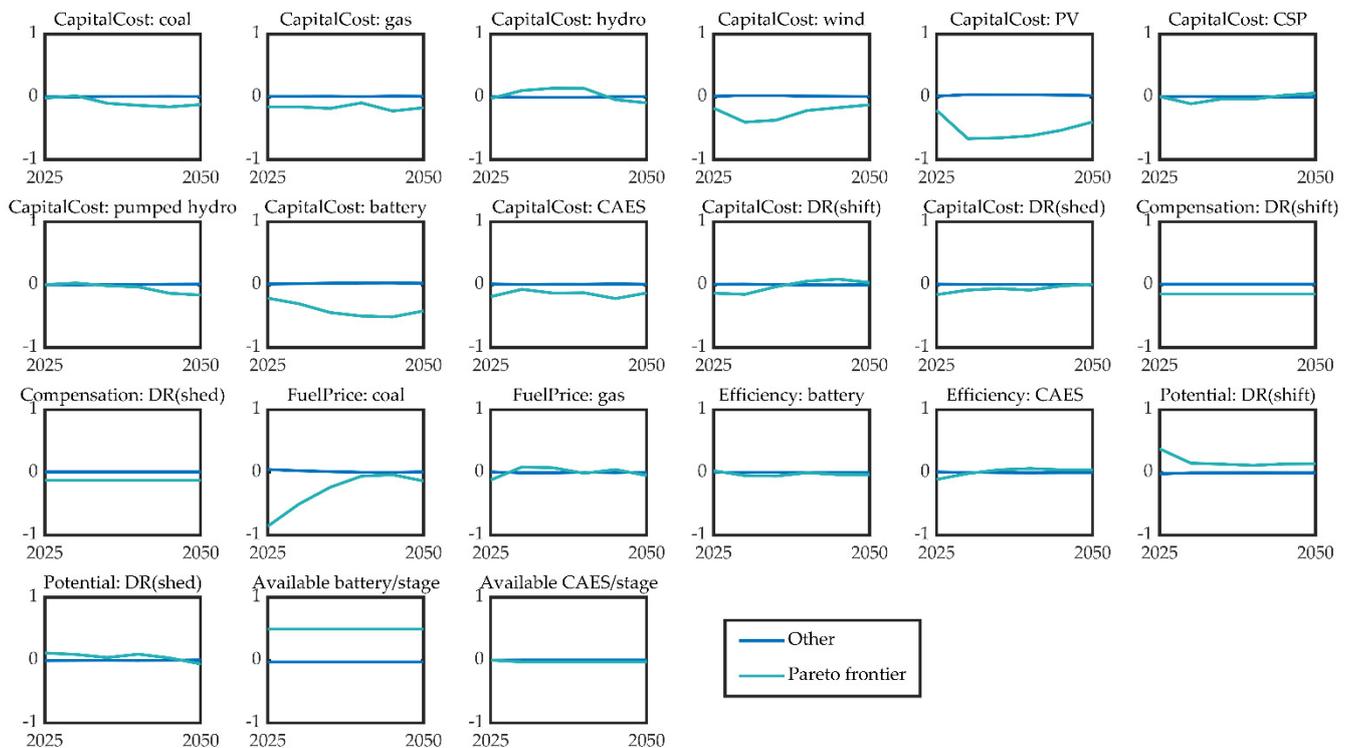


Figure 20. Normalized mean values of input parameters for optimal paths and other paths.

4. Conclusions

Power systems are transforming to incorporate high shares of VRE with the help of multiple flexibility resources. Under multiple uncertainties, we propose a data-driven approach to analyze power system transitions. Specifically, our analysis takes more than twenty uncertain factors related to generation and flexibility technologies as inputs and generates massive transition paths through the multistage planning model. The paths are visualized by the PCA method, and milestones focusing on flexibility resources are proposed. Key influential factors are identified based on the cluster technique and distance calculation. The benefits of flexibility resources are identified and calculated through clustering and the proposed marginal benefit indices. The optimal paths are identified by the Pareto frontier, and recommendations are provided.

Case studies based on the Northwest China power system show that the transition paths under multiple uncertainties are diverse and span a large range, and the traditional method deviates from the proposed method. The proposed method can consider the overall input factor space and thus is more reliable. Under the high-VRE-share target, the system LCOE shows a slight decline over time. The annual carbon emissions increase by 12% and finally decrease by 57%. The key influential factors are the capital cost of VRE generation and battery, the coal price, and the available battery investment capacity at each stage.

Flexibility resources increase much faster than the VRE penetration level. The flexibility capacity when the VRE share is 70~80% is approximately 23 times that when the share is 10~20%. Among flexibility resources, the battery will be the most important, which is likely to appear when the VRE share is 40~45%, and dominates the flexibility resources when the share is 60~65%.

Flexibility resources impact all technical, economic and economic aspects of the transition. Technically, the battery facilitates PV allocation. An increase of 1 MW in battery corresponds to approximately 1 MW in additional PV capacity. Economically, flexibility resources from both the generation and storage sides can help reduce the transition cost, and their marginal benefits are both approximately 216 k\$/MW. Furthermore, storage benefits emission reduction more than other factors, with a marginal benefit of 361 (ton/y)/MW or above.

For decision makers, the thermal unit retrofit could be encouraged in an early transition stage since it is cheaper and benefits the economy and environment. Additionally, under the early stage, the coal price should be focused on since it composes a large part of the system LCOE when the VRE share is not high. Then, policies should be proposed to encourage the development of storage because its lower cost and high capacity contributes to the optimal transition paths.

Author Contributions: Conceptualization, Z.L. and B.Z.; Funding acquisition, Z.L.; Investigation, H.L.; Methodology, H.L.; Software, H.L.; Supervision, Z.L. and B.Z.; Validation, Y.Q. and B.Z.; Writing—original draft, H.L.; Writing—review & editing, Y.Q., Z.L. and B.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Science and technology project of State Grid Corporation of China: Research on typical scenario construction and reliable operation conditions of the power system with high renewable energy penetration, Grant No. 5100-202055390A-0-0-00.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

Acknowledgments: The authors are grateful to Ningxin Ma for the suggestions and comments on the paper writing.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. The Description of the Path Generation Model

1. Objective Function

The objective is to minimize the overall cost, consisting of the investment cost, the fixed O&M cost and the variable operation cost, for the whole path:

$$\min C^{INV} + C^{OM} + C^{OPR} \quad (A1)$$

$$C^{INV} = \sum_{n=1}^N \frac{1}{(1+d)^{(n-1)Z}} \sum_{k=1}^K \sum_X a_n^X I_{k,n}^X \quad (A2)$$

$$C^{OM} = \sum_{n=1}^N \frac{1}{(1+d)^{(n-1)Z}} \frac{(1+d)^Z - 1}{d(1+d)^Z} \sum_{k=1}^K \sum_X f_n^X \bar{I}_{k,n}^X \quad (A3)$$

$$C^{OPR} = \sum_{n=1}^N \frac{1}{(1+d)^{(n-1)Z}} \frac{(1+d)^Z - 1}{d(1+d)^Z} \sum_{s=1}^S \rho_s \sum_{t=1}^T \sum_{k=1}^K \left(\begin{array}{l} \sum_G (c_n^G P_{k,n,s,t}^G + u_n^G U_{k,n,s,t}^G) \\ + \sum_{ES} c_n^{ES} (P_{k,n,s,t}^{ES,cha} + P_{k,n,s,t}^{ES,dis}) \\ + \sum_{DR} c_n^{DR} P_{k,n,s,t}^{DR} + c_n^{Cur} P_{k,n,s,t}^{Cur} \end{array} \right) \quad (A4)$$

where C^{INV} , C^{OM} and C^{OPR} represent the present value of the investment cost, fixed O&M cost and variable operation cost, respectively. d is the annual discount rate. X represents all kinds of generation and flexibility resources. a_n^X is the unit capital cost of a resource in stage n . $I_{k,n}^X$ is the investment capacity of a resource at the k th province in the n th stage. Z is the number of years per stage. f is the annual unit fixed O&M cost, and \bar{I} is the total installed capacity of a resource. There are S scenarios per stage, with each scenario representing ρ_s days in a year. c_y^G and u_y^G are the unit generation cost and start-up cost for thermal power, where $G = \{GC, GG\}$ represents the set of coal-fired and gas-fired power generation. $P_{k,n,s,t}^G$ and $U_{k,n,s,t}^G$ are the thermal generation power and start-up capacity at time t for scenario s in the n th stage, respectively. c_n^{ES} is the unit operation cost for energy storage devices, where $ES = \{PH, BT, CAES\}$ includes the pumped hydro, battery and CAES. $P_{k,n,s,t}^{ES,cha}$ and $P_{k,n,s,t}^{ES,dis}$ are the charging and discharging power, respectively. $DR = \{SHIFT, SHED\}$ includes two kinds of DR: load-shifting and load-shedding DR. c_n^{DR} and $P_{k,n,s,t}^{DR}$ are the unit compensation cost for consumers and the response power of DR, respectively. c_n^{Cur} and $P_{k,n,s,t}^{Cur}$ are the unit penalty for load curtailment and curtailment power, respectively.

2. Constraints at the First Level

The constraints at the first level include the investment decisions and the VRE generation targets. The investment capacity is bounded by the maximum available investment at each stage and the maximum total capacity at the end year. The VRE penetration target constraint, which means that a certain fraction of the total electricity demand is supplied by VREs, is similar to that in reference [39]. In this study, we set the VRE target only for the final stage.

3. Constraints at the Second Level

The second level constrains the operation simulation of each generation and flexibility technique for each scenario. For thermal units, whose flexibility constraints include the ramping rate, the minimum stable level and the minimum on/off time, the unit grouping method in reference [46] is utilized to model the flexibility limits of the units. The units are aggregated based on similar operational characteristics. In this study, we aggregate the units in each region into two groups: coal-fired units and gas-fired units.

For wind farms, PV plants, storage devices and generation reserves, the method in reference [39] is used to formulate the constraints. The constraints of CSP are similar to those in reference [46]. The models of hydro power and DR are described in reference [12]. The load curtailment for each region should be no more than the load demand at each time.

Appendix B.

Table A1. The uncertain parameters and their range.

No.	Uncertain Parameters	Unit	2025		2030		2035		2040		2045		2050		Refs.
			LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	
1	Capital cost: coal-fired thermal	\$/kW	563	625	563	625	563	625	563	625	563	625	563	625	[47]
2	Capital cost: gas-fired thermal	\$/kW	338	375	338	375	338	375	338	375	338	375	338	375	[47]
3	Capital cost: hydro	\$/kW	1418	1424	1411	1424	1411	1424	1411	1424	1405	1424	1398	1424	[42]
4	Capital cost: wind	\$/kW	1241	1385	1098	1385	1052	1379	1006	1372	980	1366	954	1359	[42]
5	Capital cost: PV	\$/kW	876	1215	562	1241	503	1183	444	1124	405	1085	366	1045	[42]
6	Capital cost: CSP	\$/kW	4998	6397	3502	6299	3280	6096	3058	5894	2960	5769	2862	5645	[42]
7	Capital cost: pumped hydro	\$/kW	659	732	659	732	659	732	659	732	659	732	659	732	[40]
8	Capital cost: battery storage	\$/kW	700	1320	496	1200	460	1160	400	1120	320	1100	300	1032	[43]
9	Capital cost: CAES	\$/kW	3019	4523	2845	4471	2781	4371	2726	4284	2562	4026	2406	3781	[40]
10	Capital cost: load-shifting DR	\$/kW	8	11	8	11	8	11	8	11	8	11	8	11	[12]
11	Capital cost: load-shedding DR	\$/kW	10	14	10	14	10	14	10	14	10	14	10	14	[12]
12	Compensation cost: load-shifting DR	\$/MWh	35	65	35	65	35	65	35	65	35	65	35	65	[12]
13	Compensation cost: load-shedding DR	\$/MWh	45	84	45	84	45	84	45	84	45	84	45	84	[12]
14	Fuel price: coal	\$/MWh	20.4	30.6	22.4	33.5	24.8	37.2	27.0	40.5	27.0	40.5	27.0	40.5	[40]
15	Fuel price: gas	\$/MWh	54.1	81.1	58.9	88.4	65.1	97.6	72.5	108.7	72.5	108.7	72.5	108.7	[40]
16	Efficiency: battery storage	-	90.0	95.0	90.0	95.0	90.0	95.0	90.0	95.0	90.0	95.0	90.0	95.0	[40]
17	Efficiency: CAES	-	59.0	70.0	59.0	70.0	59.0	70.0	59.0	70.0	59.0	70.0	59.0	70.0	[40]
18	Potential: load-shifting DR	-	0.00	0.05	0.00	0.05	0.00	0.05	0.00	0.05	0.00	0.05	0.00	0.05	
19	Potential: load-shedding DR	-	0.00	0.05	0.00	0.05	0.00	0.05	0.00	0.05	0.00	0.05	0.00	0.05	
20	Available battery investment for each stage	-	Assume linear increase with time, and the increase ratio are uncertain												
21	Available CAES investment for each stage	-	Assume linear increase with time, and the increase ratio are uncertain												

LB and UB represent lower bound and upper bound, respectively.

Table A2. The contribution of the first two components of PCA at different stages.

Stage	2025		2030		2035		2040		2045		2050	
Coal	0.00	0.00	0.00	0.00	-0.02	0.06	0.04	-0.14	-0.10	0.12	-0.07	0.01
Gas	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hydro	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Wind	1.00	-0.08	0.40	0.92	0.94	-0.32	-0.99	0.08	-0.53	-0.84	-0.70	0.29
PV	0.08	1.00	-0.92	0.40	-0.34	-0.91	0.12	0.86	0.60	-0.43	0.41	0.90
CSP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pumped hydro	0.00	0.00	-0.01	0.01	0.00	-0.03	0.00	0.01	0.00	0.00	0.00	0.00
Battery	0.00	0.00	-0.01	0.01	0.01	-0.23	-0.04	0.48	0.59	-0.30	0.58	-0.28
CAES	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01	-0.01	0.11
Load-shifting DR	0.00	0.02	-0.03	0.04	0.01	-0.02	0.00	0.00	0.00	0.00	0.00	0.08
Load-shedding DR	0.00	0.00	0.00	0.00	0.00	0.03	0.00	-0.01	0.00	0.01	0.00	0.01
Coal-fired thermal retrofit	0.00	0.00	0.00	0.00	0.00	0.01	0.01	-0.03	-0.03	0.03	-0.02	0.00
Gas-fired thermal retrofit	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cumulative explanation	1.00		0.99		0.96		0.95		0.96		0.99	

The value of the contribution ranges from -1 to 1. The larger the absolute value is, the greater the contribution. The last row indicates the cumulative explanation of the two dimensions.

Table A3. The average distances of input parameters for different aspects of transition paths.

Aspect of Transition Path	Technical		Economic	Environmental
	Generation Mix	Flexibility Mix	LCOE	Carbon Emission
CapitalCost: coal	0.03	0.09	0.06	0.07
CapitalCost: gas	0.11	0.10	0.09	0.05
CapitalCost: hydro	0.10	0.08	0.09	0.08
CapitalCost: wind	0.17	0.11	0.24	0.53
CapitalCost: PV	0.36	0.37	0.19	0.32
CapitalCost: CSP	0.15	0.13	0.14	0.08
CapitalCost: pumped hydro	0.09	0.04	0.06	0.09
CapitalCost: battery	0.46	0.55	0.22	0.25
CapitalCost: CAES	0.08	0.08	0.14	0.07
CapitalCost: DR(shift)	0.07	0.10	0.08	0.08
CapitalCost: DR(shed)	0.05	0.15	0.11	0.12
Compensation: DR(shift)	0.09	0.08	0.03	0.03
Compensation: DR(shed)	0.06	0.06	0.16	0.05
FuelPrice: coal	0.21	0.18	0.96	0.56
FuelPrice: gas	0.07	0.06	0.10	0.08
Efficiency: battery	0.05	0.06	0.10	0.07
Efficiency: CAES	0.06	0.08	0.06	0.06
Potential: DR(shift)	0.11	0.18	0.10	0.10
Potential: DR(shed)	0.11	0.10	0.09	0.08
Available battery/stage	1.24	1.39	0.12	0.31
Available CAES/stage	0.06	0.11	0.03	0.06

Table A4. The comparison between the transition results with different thermal retrofit degrees.

Retrofit Degree		No Retrofit	Deep Retrofit
Capacity in 2050 (MW)	Coal-fired thermal	106,329	111,855
	Gas-fired thermal	6	6
	Hydro	46,920	46,920
	Wind	328,872	346,541
	PV	594,441	583,673
	CSP	1600	1600
	Pumped hydro	5000	5000
	Battery	370,102	323,906
	CAES	0	0
	Load-shifting DR	1087	1087
	Load-shedding DR	11,506	11,506
	Coal-fired thermal retrofit	0	51,539
	Gas-fired thermal retrofit	0	176
	Cost (billion \$)	Investment cost	561
Fixed O&M cost		10	10
Variable operation cost		320	328
Total cost		891	886

References

- Mai, T.; Sandor, D.; Wiser, R.; Schneider, T. *Renewable Electricity Futures Study. Executive Summary*; National Renewable Energy Lab. (NREL): Golden, CO, USA, 2012.
- European Climate Foundation. Roadmap 2050. Available online: <https://www.roadmap2050.eu/reports> (accessed on 5 September 2021).
- Energy Research Institute and National Development and Reform Commission. *China 2050 High Renewable Energy Penetration Scenario and Roadmap Study*; Energy Research Institute and National Development and Reform Commission: Beijing, China, 2015.
- Mohandes, B.; Moursi, M.S.E.; Hatziaargyriou, N.; Khatib, S.E. A Review of Power System Flexibility With High Penetration of Renewables. *IEEE Trans. Power Syst.* **2019**, *34*, 3140–3155. [[CrossRef](#)]
- Alizadeh, M.; Moghaddam, M.P.; Amjady, N.; Siano, P.; Sheikh-El-Eslami, M. Flexibility in future power systems with high renewable penetration: A review. *Renew. Sustain. Energy Rev.* **2016**, *57*, 1186–1193. [[CrossRef](#)]
- International Energy Agency (IEA). *Status of Power System Transformation 2019*; IEA: Paris, France, 2019.
- International Energy Agency (IEA). *World Energy Outlook 2018*; IEA: Paris, France, 2018.
- International Energy Agency (IEA). *China Power System Transformation-Assessing the Benefit of Optimised Operations and Advanced Flexibility Options*; IEA: Paris, France, 2019.
- International Renewable Energy Agency (IRENA). *Power System Flexibility for the Energy Transition*; IRENA: Abu Dhabi, United Arab Emirates, 2018.
- Hansen, K.; Breyer, C.; Lund, H. Status and perspectives on 100% renewable energy systems. *Energy* **2019**, *175*, 471–480. [[CrossRef](#)]
- IRENA. *Innovation Landscape for a Renewable-Powered Future*; IRENA: Abu Dhabi, United Arab Emirates, 2019.
- Li, H.; Lu, Z.; Qiao, Y.; Zhang, B.; Lin, Y. The flexibility test system for studies of variable renewable energy resources. *IEEE Trans. Power Syst.* **2021**, *36*, 1526–1536. [[CrossRef](#)]
- Venkataraman, S.; Jordan, G.; O'Connor, M.; Kumar, N.; Lefton, S.; Lew, D.; Brinkman, G.; Palchak, D.; Cochran, J. *Cost-Benefit Analysis of Flexibility Retrofits for Coal and Gas-Fueled Power Plants: August 2012–December 2013*; National Renewable Energy Lab.(NREL): Golden, CO, USA, 2013.
- Connolly, D.; Lund, H.; Mathiesen, B.V.; Leahy, M. The first step towards a 100% renewable energy-system for Ireland. *Appl. Energy* **2011**, *88*, 502–507. [[CrossRef](#)]
- Dominković, D.F.; Bačeković, I.; Čosić, B.; Krajačić, G.; Pukšec, T.; Duić, N.; Markovska, N. Zero carbon energy system of South East Europe in 2050. *Appl. Energy* **2016**, *184*, 1517–1528. [[CrossRef](#)]
- Mathiesen, B.V.; Lund, H.; Karlsson, K. 100% Renewable energy systems, climate mitigation and economic growth. *Appl. Energy* **2011**, *88*, 488–501. [[CrossRef](#)]
- Panos, E.; Kober, T.; Wokaun, A. Long term evaluation of electric storage technologies vs alternative flexibility options for the Swiss energy system. *Appl. Energy* **2019**, *252*, 113470. [[CrossRef](#)]
- Morgan, M.G.; Keith, D.W. Improving the way we think about projecting future energy use and emissions of carbon dioxide. *Clim. Chang.* **2008**, *90*, 189–215. [[CrossRef](#)]
- Syri, S.; Lehtilä, A.; Ekholm, T.; Savolainen, I.; Holttinen, H.; Peltola, E. Global energy and emissions scenarios for effective climate change mitigation—Deterministic and stochastic scenarios with the TIAM model. *Int. J. Greenh. Gas Control* **2008**, *2*, 274–285. [[CrossRef](#)]
- Bistline, J.E.; Weyant, J.P. Electric sector investments under technological and policy-related uncertainties: A stochastic programming approach. *Clim. Chang.* **2013**, *121*, 143–160. [[CrossRef](#)]
- Usher, W.; Strachan, N. Critical mid-term uncertainties in long-term decarbonisation pathways. *Energy Policy* **2012**, *41*, 433–444. [[CrossRef](#)]
- Labriet, M.; Nicolas, C.; Tchung-Ming, S.; Kanudia, A.; Loulou, R. Energy decisions in an uncertain climate and technology outlook: How stochastic and robust methodologies can assist policy-makers. In *Informing Energy and Climate Policies Using Energy Systems Models*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 69–91.
- Babonneau, F.; Kanudia, A.; Labriet, M.; Loulou, R.; Vial, J.-P. Energy security: A robust optimization approach to design a robust European energy supply via TIAM-WORLD. *Environ. Modeling Assess.* **2012**, *17*, 19–37. [[CrossRef](#)]
- Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. *Global Sensitivity Analysis: The Primer*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
- Kwakkel, J.H.; Pruyt, E. Exploratory modeling and analysis, an approach for model-based foresight under deep uncertainty. *Technol. Forecast. Soc. Chang.* **2013**, *80*, 419–431. [[CrossRef](#)]
- Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006.
- Pye, S.; Sabio, N.; Strachan, N. An integrated systematic analysis of uncertainties in UK energy transition pathways. *Energy Policy* **2015**, *87*, 673–684. [[CrossRef](#)]
- Moallemi, E.A.; de Haan, F.; Kwakkel, J.; Aye, L. Narrative-informed exploratory analysis of energy transition pathways: A case study of India's electricity sector. *Energy Policy* **2017**, *110*, 271–287. [[CrossRef](#)]
- Yue, X.; Patankar, N.; Decarolis, J.; Chiodi, A.; Rogan, F.; Deane, J.P.; O'Gallachoir, B. Least cost energy system pathways towards 100% renewable energy in Ireland by 2050. *Energy* **2020**, *207*, 118264. [[CrossRef](#)]
- Bosetti, V.; Marangoni, G.; Boronovo, E.; Diaz Anadon, L.; Barron, R.; McJeon, H.C.; Politis, S.; Friley, P. Sensitivity to energy technology costs: A multi-model comparison analysis. *Energy Policy* **2015**, *80*, 244–263. [[CrossRef](#)]

31. Zsiborács, H.; Baranyai, N.H.; Vincze, A.; Zentkó, L.; Birkner, Z.; Máté, K.; Pintér, G. Intermittent renewable energy sources: The role of energy storage in the European power system of 2040. *Electronics* **2019**, *8*, 729. [[CrossRef](#)]
32. Sadiqa, A.; Gulagi, A.; Breyer, C. Energy transition roadmap towards 100% renewable energy and role of storage technologies for Pakistan by 2050. *Energy* **2018**, *147*, 518–533. [[CrossRef](#)]
33. Cebulla, F.; Naegler, T.; Pohl, M. Electrical energy storage in highly renewable European energy systems: Capacity requirements, spatial distribution, and storage dispatch. *J. Energy Storage* **2017**, *14*, 211–223. [[CrossRef](#)]
34. Liu, H.; Brown, T.; Andresen, G.B.; Schlachtberger, D.P.; Greiner, M. The role of hydro power, storage and transmission in the decarbonization of the Chinese power system. *Appl. Energy* **2019**, *239*, 1308–1321. [[CrossRef](#)]
35. China National Renewable Energy Center (CNREC). *China Renewable Energy Outlook 2018*; CNREC: Beijing, China, 2018.
36. Mongird, K.; Fotedar, V.; Viswanathan, V.; Koritarov, V.; Balducci, P.; Hadjerioua, B.; Alam, J. *Energy Storage Technology and Cost Characterization Report*; Pacific Northwest National Lab.(PNNL): Richland, WA, USA, 2019.
37. Gils, H.C. Assessment of the theoretical demand response potential in Europe. *Energy* **2014**, *67*, 1–18. [[CrossRef](#)]
38. Yue, X.; Pye, S.; DeCarolis, J.; Li, F.G.N.; Rogan, F.; Gallachóir, B.Ó. A review of approaches to uncertainty assessment in energy system optimization models. *Energy Strategy Rev.* **2018**, *21*, 204–217. [[CrossRef](#)]
39. Chen, X.Y.; Lv, J.J.; McElroy, M.B.; Han, X.N.; Nielsen, C.P.; Wen, J.Y. Power system capacity expansion under higher penetration of renewables considering flexibility constraints and low carbon policies. *IEEE Trans. Power Syst.* **2018**, *33*, 6240–6253. [[CrossRef](#)]
40. Bogdanov, D.; Farfan, J.; Sadovskaia, K.; Aghahosseini, A.; Child, M.; Gulagi, A.; Oyewo, A.S.; de Souza Noel Simas Barbosa, L.; Breyer, C. Radical transformation pathway towards sustainable electricity via evolutionary steps. *Nat. Commun.* **2019**, *10*, 1077. [[CrossRef](#)]
41. Moallemi, E.A.; Köhler, J. Coping with uncertainties of sustainability transitions using exploratory modelling: The case of the MATISSE model and the UK's mobility sector. *Environ. Innov. Soc. Transit.* **2019**, *33*, 61–83. [[CrossRef](#)]
42. Tsiropoulos, I.; Tarvydas, D.; Zucker, A. *Cost Development of Low Carbon Energy Technologies-Scenario-Based Cost Trajectories to 2050, 2017 ed.*; European Union: Maastricht, The Netherlands, 2018.
43. Cole, W.J.; Frazier, A. *Cost Projections for Utility-Scale Battery Storage*; National Renewable Energy Lab. (NREL): Golden, CO, USA, 2019.
44. Global Energy Interconnection Development and Cooperation Organization (GEIDCO). *Study on China's 14th Five-Year Development Plan for the Power Sector*; GEIDCO: Bangkok, Thailand, 2020.
45. Zhang, N.; Dai, H.; Hu, Z. A Source-Grid-Load Coordinated Planning Model Considering System Flexibility Constraints and Demand Response. *Electr. Power* **2019**, *52*, 61–69. [[CrossRef](#)]
46. Du, E.; Zhang, N.; Hodge, B.-M.; Wang, Q.; Kang, C.; Kroposki, B.; Xia, Q. The role of concentrating solar power toward high renewable energy penetrated power systems. *IEEE Trans. Power Syst.* **2018**, *33*, 6630–6641. [[CrossRef](#)]
47. China Electric Power Planning & Engineering Institute. *Report on China's Electric Power Development 2020*; People's Daily Press: Beijing, China, 2021.