

## Research article

## Evolution of renewable energy laws and policies in China

Wenyu Ma<sup>a</sup>, Wenyu Wang<sup>b,\*</sup><sup>a</sup> SINOPEC Geophysical Research Institute Co. Ltd, Nanjing City, China<sup>b</sup> City University of Hong Kong, Hong Kong, China

## ARTICLE INFO

Dataset link: <https://www.pkulaw.com>

## Keywords:

LDA  
Topic modelling  
Renewable energy policy  
Word cloud  
Perplexity

## ABSTRACT

This study employs Latent Dirichlet Allocation (LDA) topic modelling methodology to analyze documents related to renewable energy laws and policies at the central level in China. The objective is to investigate the development and evolution of renewable energy policies in China and to gain insights into the national-level attitudes towards renewable energy development. The study consists of two phases: initially, renewable energy policy documents undergo keyword analysis using word clouds and keyword co-occurrence network analysis to elucidate the focal areas and their interconnections within the legal and policy texts. Subsequently, after determining the optimal number of topics for modelling based on topic perplexity and consistency results, the text undergoes data cleaning to isolate words with practical significance. These words are then incorporated into the LDA topic model to analyze the distribution and content of potential topics within the policies. Lastly, by linearly segmenting the time frame, changes in topic intensity over time are visually examined using heat maps. The findings indicate that energy policies have consistently prioritized “development” and emphasized the significance of “new energy” in renewable energy policies. Moreover, as renewable energy has progressed, governments and policymakers have come to acknowledge the importance of comprehensive energy planning, transitioning to clean energy sources, and regulating the electricity market. This growing awareness has led to efforts to strengthen policy and regulatory measures to foster renewable energy’s sustainable development and utilization. In summary, this study highlights the effectiveness of the LDA topic model in analyzing renewable energy policies, advancing its adoption and furthering research in the field.

## 1. Introduction

As the global population continues to grow and economic development accelerates, climate change and energy demand challenges have become a matter of global concern. Managing the immense energy demand while adhering to ecological constraints presents a primary challenge for energy policies [1]. Among these policies, clean energy policy emerges as a critical strategy to address these pressing challenges affecting sustainable development [2]. Facing significant environmental pollution and energy supply challenges, China, one of the world’s leading carbon emitters, has implemented various policy measures and initiatives. These encompass promoting the Renewable Energy Law, the Clean Energy Development Plan, and establishing a carbon emissions trading system. Additionally, China set “Dual Decarbonization Goals” in 2020, aimed at advancing towards a more sustainable and competitive energy mix and fostering the consolidation and enhancement of related energy policies. While these policies have benefited China, they have also profoundly impacted the global community [3].

\* Corresponding author.

E-mail addresses: [mawy.swty@sinopec.com](mailto:mawy.swty@sinopec.com) (W. Ma), [werren01@163.com](mailto:werren01@163.com) (W. Wang).<https://doi.org/10.1016/j.heliyon.2024.e29712>

Received 8 November 2023; Received in revised form 3 April 2024; Accepted 14 April 2024

Available online 18 April 2024

2405-8440/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Among the above measures, policies targeting renewable energy are even more significant. These policies aim to promote renewable energy technology innovation, promote the development of related industries, and accelerate the popularisation of renewable energy applications. Their significance lies not only in mitigating climate change and protecting the environment but is also crucial to ensuring the security of energy supply and promoting sustainable economic development. Hence, there is a need for further in-depth studies and analyses of the evolution and effectiveness of China's renewable energy policies, as well as identifying their future trends and challenges. Several studies have already delved into the development of renewable energy policies in China. For instance, Song et al. examined the current state of China's renewable energy development alongside the challenges and remedies of related policy instruments, employing literature analysis and energy consumption market analysis [4]. Wei et al. utilised a dynamic general equilibrium model to assess the efficacy of renewable energy policies in reshaping China's low-carbon energy framework [5]. Additionally, Song et al. scrutinised the deficiencies and enhancement pathways of China's renewable energy subsidy system, stemming from the current landscape of China's renewable energy industry [6]. These studies offer insights into the development and progression of China's renewable energy policy and its impact from both qualitative and quantitative angles, rendering them highly significant. Nonetheless, the research methodology often remains detached from the contemporary policy text milieu, limiting the comprehensive analysis of policy text evolution. Hence, it is necessary to employ topic modelling in scrutinising China's renewable energy policy texts to achieve a more holistic understanding of the evolution of these policies.

In order to remedy the shortcomings mentioned above, this study adopts a two-step approach. First, renewable energy policies were keyword analysed using word cloud and keyword co-occurrence network analysis to elucidate the focus areas and their linkages in the legal and policy texts. Secondly, data cleaning was used to disentangle words of practical significance from the text in the policy and legal documents and introduce them into Latent Dirichlet Allocation (LDA) topic modelling to analyse the distribution and content of potential topics in the policy and to disentangle the time window to analyse changes in topic strength over time. The ultimate goal is to explore the intricate relationship between China's renewable energy policies and their development and to summarise the trajectory and evolution of renewable energy policies by identifying trends in themes and keywords using LDA topic modelling. Overall, the results of this study will significantly deepen the understanding of the evolution of renewable energy policies and provide important insights and recommendations to guide policy formulation and implementation. In addition, the findings will illustrate the effectiveness and multifaceted application potential of LDA topic modelling in renewable energy policy analysis, thereby promoting its more comprehensive application in renewable energy policy and helping policymakers and stakeholders study and understand it in greater depth.

## 2. Literature review

Developing renewable energy policies is a key focus area for sustainable economic growth and environmental protection. On the global stage, all nations have embraced the principle of state-backed policy support to drive renewable energy development, recognising that governmental backing is integral to its progress. However, the specifics of these policies vary across countries. In the United States, for instance, a robust policy framework is essential for promoting renewable energy and enhancing energy efficiency, although nascent energy technologies might be undervalued. In contrast, China's approach to renewable energy policy diverges from the market-driven model of Western nations, instead employing a top-down command-and-control strategy. Nevertheless, this approach has proven effective in rapidly disseminating renewable energy technologies.

At the same time, China's renewable energy policy is multi-faceted and has evolved alongside renewable energy development over time. The promulgation of the Renewable Energy Law in 2005 marked the formal inclusion of renewable energy within China's legally regulated development framework. Subsequent amendments to the law in 2009 reflected the maturation of regulated renewable energy development in China, with the scope of adjustment gradually being finalised. Following the implementation of two national five-year plans, China's renewable energy development achieved significant scale, witnessing comprehensive growth across various emerging energy sources and entering a phase of rapid expansion. In response to the global transformation of energy systems, China has further constructed a new type of energy development and standardisation system under the guidance of The Dual Decarbonization Goals, making renewable energy the primary development path. Concurrently, renewable energy policies have undergone drastic changes to adapt to the rapid energy development. Li and Taihag find evidence of a shift in the policy mix from a limited number of authority-based instruments to a broader range of intensive instruments with diverse types [7]. The Chinese government has increased its policy efforts to reduce air pollution, altered its support for renewable energy, and employed innovative policy tools to mitigate CO<sub>2</sub> emissions. Additionally, Yang and Lo employ a systematic review approach to analyse primary renewable energy and energy efficiency policies executed across five sectors in China: power, industry, transportation, buildings, and local government [8]. The study identifies the challenges these policies face and offers policy recommendations. The complexity of this policy landscape poses unprecedented challenges for traditional policy analysis and necessitates more significant efforts from policymakers and stakeholders to comprehend policy trends. However, the current understanding of the evolution of China's renewable energy policies has mainly relied on systematic policy analysis and textual analysis, with limited use of topic modelling and keyword analysis in the context of renewable energy policies.

In recent years, the rapid development of artificial intelligence (AI) technology has profoundly impacted various aspects of people's lives and production. A study by Yu and Xian, which utilised LDA to analyse artificial intelligence research topics, indicates an upward trend in machine learning and natural language processing within the AI field [9]. Topic modelling is a powerful text mining technique that integrates natural language processing (NLP) and information retrieval techniques to achieve data generalisation by extracting latent variables corresponding to potential topics or themes and representing documents as mixtures of probabilistic topics and words. Among various approaches to topic modelling, LDA stands out as one of the most popular methods in this field [10]. LDA

as an NLP tool, is a Bayesian network that analyses large amounts of unclassified textual data [11]. By calculating the co-occurrence frequency of different words, LDA can recognise topics in large corpora [12], which can be used in a wide range of applications from document modelling to computer vision [13].

First and foremost, LDA is an unsupervised machine learning technique for discovering hidden topics in large-scale document collections [14]. The principle lies in the fact that using LDA, topics in the corpus can be automatically generated, words in the corpus can be segmented based on these topics [15], and more meaningful topics can be extracted by calculating the probability of each topic appearing in a document, thus reflecting the importance of the topic in the document [16]. In addition, combined with visualisation techniques, it enables a more in-depth evaluation of topics and probability ratios, which helps to understand policy development and optimisation trends.

Second, LDA has some advantages in topic modelling. As Hagen observed, the topics generated by LDA match well with independent content analysis and are less susceptible to human bias [17]. Meanwhile, Yokomoto et al. demonstrated that LDA-based topic modelling provides a quick overview of the content distribution in search results [18]. However, not all estimated themes are equally important, and not all correspond to the underlying themes in the domain. Manual checking is required to identify meaningful topics.

Third, LDA is currently widely used in the field of policy analysis. For example, the studies of Shiryaev et al. and Yue et al. used LDA models to analyse policy documents and identify trends in technical knowledge [19,20]. In addition, Zhang et al. proposed an integrated model that uses LDA to identify similarities between multiple policy texts to help government agencies develop and optimise appropriate economic policies [21]. These papers demonstrate the utility of LDA models in identifying trends in the policy domain. Meanwhile, with the rapid development of data resources, quantitative methods can identify research frontiers, providing more perspectives and strategies for textual analysis of policies [22].

In summary, advancing China's renewable energy sector has coincided with establishing a more comprehensive framework of pertinent laws and institutions at the central level. This framework has served as a crucial foundation for the subsequent development of local institutions and policies. With the implementation of China's phased transition plan and dual-carbon objectives, the significance of the renewable energy system has become increasingly prominent, and the changes manifested at the legislative level have become more apparent. However, current scholarly research on China's renewable energy system primarily focuses on literature and empirical analysis, somewhat detached from the policy's textual context. This detachment makes it challenging to interpret and analyze the relevant policies and institutional documents comprehensively. Leveraging the potent application of LDA topic modelling in textual analysis, examining Chinese renewable energy policy texts will deepen the understanding of the evolution of Chinese renewable energy policies. Furthermore, it will offer valuable insights and recommendations to inform policy formulation and implementation.

### 3. Methodology

This study analyzes the evolution of China's renewable energy policies using LDA topic modelling approach, which proposes an innovative multidimensional research method for policy making in the field of renewable energy.

The operation process includes the following steps, as shown in Fig. 1: first, the policy documents are preprocessed. Use Python tools to download and extract critical information, such as content, title, and time, from the text of China's renewable energy laws and policies. In this process, noise data such as duplicate documents, documents unrelated to the topic and documents lacking critical information are artificially eliminated. The results are organized into the required data forms. Split Chinese text into words using "jieba". "jieba" is a popular Chinese word-splitting tool based on an algorithm with a prefix lexicon and statistical methods to identify words in Chinese text and split them into words as data input for subsequent data analysis and modelling. Subsequently, stop word lists and custom dictionaries were used for further cleaning and optimization to ensure the dataset was suitable for subsequent topic modelling analysis.

Next, the policy text was subjected to 'keyword analysis'. The preprocessed data was analyzed using word clouds and keyword co-occurrence network analysis. The word cloud outputs high-frequency word results by counting the frequency of words in the renewable energy law and policy texts to extract the critical information of the said texts quickly. Based on the obtained high-frequency keywords, the co-occurrence matrix of the words is constructed according to the research purpose. Based on the co-occurrence matrix, the co-occurrence network graph of the keywords is constructed, which visualizes and reveals the correlation relationship between the keywords in the policy texts, where the size of nodes in the visual graph indicates the frequency of occurrence of the keywords, the colour depth of the nodes indicates the importance of the keywords, and the thickness of the connecting lines indicates the strength of the keywords occurring at the same time. Through the analysis of the keyword co-occurrence network, the core keywords and their connections with the Chinese renewable energy law and policy texts were derived, which were used to elucidate the focus areas and their connections in the law and policy texts.

Finally, words with practical significance separated from the policy texts after the above data preprocessing were introduced into LDA topic modelling. The optimal number of topics was determined using a combination of topic perplexity and consistency. The LDA model in the Gensim library was used for topic modelling and to obtain the optimal topic distribution. By analyzing the output of the LDA model to understand the relationship between the words and topics of interest for each topic, a deeper understanding of the content of renewable energy laws and policies is obtained. In addition, by partitioning the policies in linear time, this study analyzes the change in topic strength over time by calculating the topic strength for each time window using a visual heat map.

The data used in this paper is sourced from the legal normative texts on China's renewable energy policies available in Peking University's Fabao Policy Database. The selection primarily focuses on laws and departmental regulations at the central level, highlighting the analysis of binding laws and policy texts. The dataset comprises 158 policy and legal texts from 1995 to 2023, excluding non-text documents, resulting in a maximum sample size of 157.

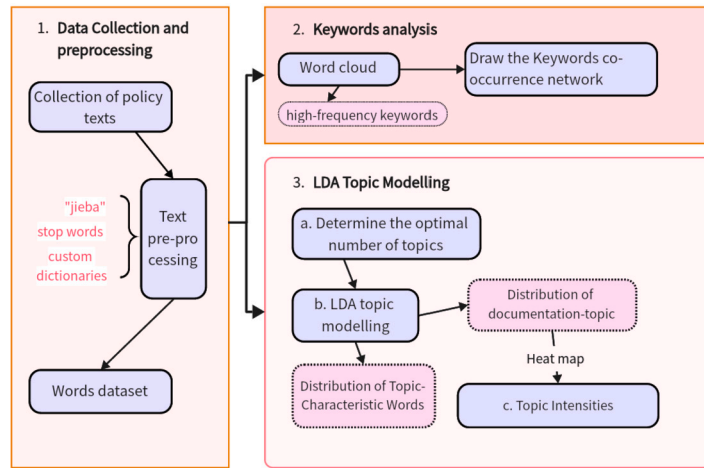


Fig. 1. Research framework diagram.

Table 1  
High-frequency words in policy texts (top 20).

High-frequency vocabulary	Word frequency/time	High-frequency vocabulary	Word frequency/time
Projects	2816	Enterprises	971
Development	1677	Engineering	965
Company limited	1655	Funding	916
Wind Power	1626	Construction	911
Power Generation	1445	Wind Power	908
Wind Farms	1313	Subsidies	875
Technology	1292	Electricity	870
Energy	1037	State	833
Power Generation Engineering	1010	Company	829
Danwei	1000	Tariffs	824

#### 4. Data analysis result and discussion

##### 4.1. Policy keyword analysis

###### 4.1.1. High frequency vocabulary analysis

This study begins by preprocessing the collected renewable energy policy texts. The goal is to generate a cleaned and optimised dataset for subsequent topic modelling and classification tasks. The Python Word Cloud library is then used to process the texts, selecting high-frequency words to generate word cloud images and outputting the top 20 high-frequency nouns, as detailed in Table 1.

The Word Cloud results visually depict the frequency and importance of keywords in renewable energy policy texts. First and foremost, it is evident that the “Development” of renewable energy has consistently been the focal point of energy policy. Similarly, the high-frequency words “Project,” “Engineering,” and “Construction” validate the notion that policy orientation drives the rapid development of related industries. Additionally, high-frequency terms like “Company,” “Enterprise,” and “Danwei” are concentrated, emphasising the significance of these entities in renewable energy policy. It reveals that the current focus of renewable energy policy or legislation is on these subjects. In other words, these entities represent the most active groups within renewable energy development and policy adjustment. Furthermore, high-frequency terms such as “Wind Power,” “Power generation,” “Power Generation Project,” and “Technology” indicate that the current utilisation of renewable energy is primarily concentrated in the power generation sector. Policies and laws have also emphasised renewable energy in this field, underscoring its potential in the electricity industry. Finally, the inclusion of high-frequency words like “Funds,” “Subsidies,” and “Tariffs” suggests that current renewable energy policies are also geared towards making adjustments from a market perspective to foster the development of renewable energy.

In summary, analysing renewable energy policies using Word Cloud can identify high-frequency words in these policies, which can help renewable energy policymakers and researchers better understand the core content and concerns of policy texts. These insights are essential for policymakers to develop and refine policies and guide decision-making.

###### 4.1.2. Keyword co-occurrence network analysis

The previous High Frequency Vocabulary Analysis allows us to identify the keywords with the highest frequency of occurrence in the policy documents. This study further carries out keyword co-occurrence network analysis to investigate these keywords’ relevance and co-occurrence patterns. Based on the previously mentioned data preprocessing, this study continues to use the bag-

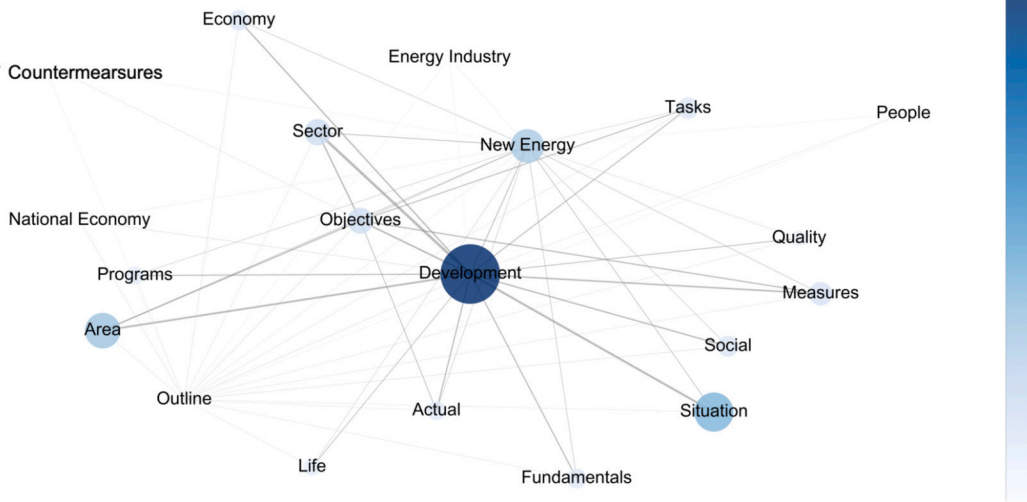


Fig. 2. Keyword Co-occurrence Network.

of-words model to convert the text in the dataset into a vector representation. Subsequently, the keyword co-occurrence network is constructed based on the co-occurrence relationship between keywords. The specific results are shown in the following Fig. 2.

In the keyword co-occurrence network, nodes are selected for the higher-frequency feature words according to the established criteria in the High Frequency Vocabulary Analysis. Additionally, words irrelevant to policy or legal norms are excluded from the network to ensure its adjustment under the semantic relationship among the generated keywords. The edges connecting the nodes represent the co-occurrence probability, which is calculated by concurrently determining the number of texts containing both nodes. The size of the nodes corresponds to the frequency of the keywords in the co-occurrence matrix. The shade of their colour indicates their importance, and the thickness of the connecting lines reflects the intensity of the co-occurrence between the keywords. By visualizing the results of the keyword co-occurrence network, we can observe the core keywords and their relevance.

Notably, the keyword “Development” holds a central position in policy research and regulatory adjustments. It ranks among the top twenty core words in frequency within the keyword co-occurrence network, displaying the darkest colour and occupying the most significant area proportion. This result indicates a high likelihood of the word “Development” recurring in policy texts, which is consistent with the findings of the high-frequency word analysis. Notably, the thick connecting lines indicate a higher probability of co-occurrence between “Development” and words such as “Society,” “Area,” and “Environment,” emphasizing the significance of considering regional disparities and practical aspects in renewable energy development. Moreover, the development of energy not only contributes to “Society” and “Area” but also demonstrates a close relationship with the “Environment.” Additional keywords such as “measures,” “tasks,” “sectors,” “Objectives,” and “Programs” emerge as crucial considerations in the realm of renewable energy development. The interplay between “Development” and “Programs” reflects the prevailing form of current renewable energy policy, encompassing a comprehensive normative framework. This framework guides the industry by outlining sectoral responsibilities, establishing development goals and plans, and implementing practical tasks and measures. Furthermore, the co-occurrence of “Economy,” “National Economy,” “People,” and “Life” emphasizes the importance of renewable energy. It indicates that renewable energy policies are oriented towards market economics, prioritize economic development, and revolve around the well-being of individuals.

In addition, “New Energy” assumes a central position alongside “development” within the keyword co-occurrence network, signifying its significance within policy texts related to renewable energy. Although the co-occurrence probability between “Development” and other high-frequency keywords may not be as high as with “New Energy,” the results still indicate a relatively comprehensive normative framework for renewable energy development within the policy text. Specific nodes, such as “Energy Industry,” exhibit a lower probability of co-occurrence with other nodes. However, it is essential to note that despite their lower co-occurrence probabilities, these individual nodes may still have a higher likelihood of co-occurrence with other potential keywords within the actual policy text. Therefore, it is necessary to continue completing the LDA modelling to uncover potential themes and content.

In conclusion, by integrating the results from the high-frequency word analysis and the keyword co-occurrence network analysis, a more comprehensive understanding of the characteristics of keywords in policy documents and the focal points of policies can be achieved. This analytical approach facilitates a deeper comprehension of the keywords present in policy documents and their areas of emphasis.

## 4.2. Topic clustering results

### 4.2.1. Number of topics identified

In this paper, we utilize topic perplexity and topic consistency to determine the best number of topics. When selecting the optimal number of topics, observing the trends of topic perplexity and consistency is essential as the number of topics varies. Generally, perplexity decreases as the number of topics increases, but if the number of topics becomes too large, the model may need to be more

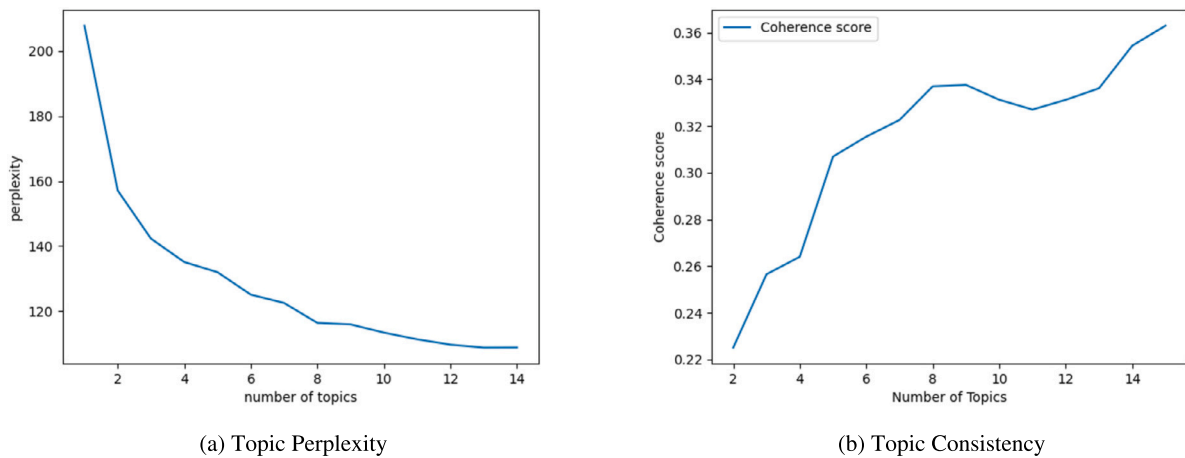


Fig. 3. Topic Perplexity and Coherence Score.

**Table 2**  
Distribution of topics in renewable energy policies.

Topic type	Topic content
Topic 0	Specialized Funding and Policy Implementation
Topic 1	Research and application of ocean energy technologies
Topic 2	Renewable energy policies in the building sector
Topic 3	Planning and utilization of renewable energy technologies
Topic 4	National energy planning and market regulation
Topic 5	Wind power and new energy generation projects for renewable energy companies
Topic 6	Wind power and biomass power generation engineering and technology
Topic 7	Renewable energy subsidy policies and tariff surveys

balanced. Thus, the critical point at which perplexity begins to decrease must be identified as the optimal number of topics. In our study, we have set the adjustable range of the number of topics to be 2-15. We evaluate this range using the Log Perplexity method in the LDA model and compare the logarithmic perplexity between different models to assess the model's fit. The figure below presents the specific results. From Fig. 3a, it can be observed that the decrease in perplexity slows down when the number of topics exceeds 8.

Similarly, after determining the approximate range of topics using perplexity, we employ the topic consistency score to determine the optimal number of topics. Topic consistency measures the relevance and differentiation of words within a topic. A high consistency score indicates strong correlation and differentiation within a topic, while a low consistency score suggests weak relationships or confusion with words from other topics. In this paper, based on the data processing described earlier, we construct a topic model using LDA. Once the topic model is obtained, we calculate the consistency scores for each topic, which are presented in Fig. 3b below. In this figure, the topic consistency score slows when the number of topics reaches 8.

To summarize, the results of topic perplexity can range in the number of topics above 8. However, given that the number of topics selected for perplexity tends to be large, this can lead to a large degree of similarity between the topics extracted and a low degree of topic recognition. The topic consistency score obtained in the second step measures the interpretability of the different numbers of topics. The results show that the number of topics at 8 can make the interpretability acceptable. Also, by comparing the results of keywords with different numbers of topics, it can be found that the repetition of keywords between topics is low when the number of topics is 8. Therefore, we chose 8 as this paper's optimal number of topics.

#### 4.2.2. Analyzing topic content

After analyzing the high-frequency vocabulary and co-occurrence networks, we used LDA topic modelling to generate topic-vocabulary and document-topic matrices with the aim of discovering potential topics in the renewable energy policy text corpus and describing the relationships between different topics, specific vocabularies, and documents. Table 2 lists the simplified results, providing the content of each topic, and Table A.1 lists the detailed results, providing the top 6 feature words for each topic and their relative proportions to the total weight of the feature words within the respective topic.

Topic 0 deals with earmarked funding and policy implementation, including management and allocation of funds, expert studies, monitoring and evaluation reports. Policies under this theme aim to provide economic support, policy guidance and market incentives to promote renewable energy development. The objective is to promote innovation, deployment and diffusion of renewable energy technologies.

Topic 1 focuses on the research and application of marine energy technologies, covering aspects of marine energy engineering, equipment design, testing and declaration. As a clean and renewable energy source, ocean energy offers diversified options in renewable energy. Policies encouraging the development and utilisation of ocean energy technologies can help reduce reliance on traditional energy sources, thereby promoting the transformation of the energy mix and fostering sustainable development.

Topic 2 revolves around renewable energy policies in the building sector, including energy-saving technologies for buildings, solar energy utilisation and their application in urban and rural housing. The building sector is an essential area of energy consumption, and policies prioritising energy-efficient building technologies can help reduce energy consumption. In addition, renewable energy policies targeting solar energy utilisation in buildings could achieve self-sufficiency in energy supply, promote green buildings and sustainable housing, improve the living environment for urban and rural residents, and ultimately contribute to sustainable urban development.

Topic 3 is the planning and utilisation of renewable energy technologies, including solar energy, wind energy, biomass energy resource utilisation and industrial development in rural areas. The planning and utilisation of renewable energy technologies promoted the diversification and sustainable use of energy resources. Renewable energy sources, such as solar and wind, provide an inexhaustible supply of energy, while biomass can be continuously replenished through recycling. Utilising these energy sources helps prevent resource depletion, thereby contributing to the stability and sustainability of energy supplies.

Topic 4 covers national energy planning and market regulation, including annual electricity plans, trading rules, social supervision, and the responsibilities and targets of power companies. Policies can guide and incentivize power companies to increase the capacity and proportion of renewable energy generation. Promote renewable energy development and deployment through strategic planning and target setting.

Topic 5 focuses on wind and new energy generation projects undertaken by renewable energy companies, covering large-scale companies and detailed design aspects of wind farms, wind and bioenergy generation projects. The existence and growth of renewable energy companies not only energizes the energy sector but also contributes to developing a green economy. These companies play a crucial role in promoting the innovation and application of renewable energy technologies and driving the development of related industries.

Topic 6 focuses on wind power and biomass power engineering and technology, especially in solar power, biomass power, and renewable energy utilization. As technology advances, the economic barriers to renewable power generation are reduced, thus making renewable energy an increasingly competitive option.

Theme 7 focuses on subsidy policies and tariff surveys for renewable energy, including government funding, subsidy measures, tariff standards and user quotas. Government funding and subsidies can effectively mitigate the costs associated with the renewable energy industry and improve market competitiveness. In addition, setting tariff standards and user quotas increases the market demand for renewable energy and incentivizes the consumption and utilization of renewable energy.

#### 4.3. Analyzing the evolution of thematic intensity

Based on the LDA topic modelling results discussed earlier, this section creates different time windows based on the previously mentioned overview of renewable energy policies. It assigns each theme to the corresponding time window. Then, the intensity of each theme within each time window is calculated by normalizing the theme counts, and a heat map is used to visualize the change in theme intensity. The detailed analysis results are shown in Fig. 4.

According to the heat map, the focus of renewable energy policy before 2005 was mainly on Topic 3: “Planning and utilization of renewable energy technologies”, reflecting the issues and needs that renewable energy policy was designed to address at that time. Such a characterization can be attributed to the growing awareness of climate change and environmental issues during this period when policymakers considered environmental protection essential. In addition, because the development and utilization of renewable energy technologies in China were relatively new and immature, policymakers prioritized planning and infrastructure development to establish an industrial base and technical support system.

Between 2006 and 2010, renewable energy policy shifted to Topic 2: “Renewable energy policies in the building sector”. At the same time, all the previously mentioned themes in policy and law have evolved, except for Topic 6: “Wind power and biomass power generation engineering and technology”. This shift suggests that the focus of renewable energy policy has expanded from purely technology development to its utilization in different sectors. The development of these other themes suggests that policymakers may recognize the importance of diversification and decentralization of renewable energy supply. The development and policy framework for renewable energy is gradually being implemented.

From 2010 to 2015, renewable energy policy has evolved evenly across themes. Topic 2: “Renewable energy policies in the building sector”, and Topic 7: “Renewable energy subsidy policies and tariff surveys” are relatively more prominent. This balanced development suggests that the period was characterized by a relatively high level of interest in this topic. This balanced development indicates that policymakers have adopted a comprehensive development strategy during this period and that governments have taken active measures to encourage the development and promotion of renewable energy.

Topic 4: “National energy planning and market regulation” is the focus of renewable energy policy in 2016-2023, significantly more than any other theme. Notably, between 2016 and 2020, the policy weight of all themes except Theme 4 began to decline. Subsequently, after 2020, Themes 5, 6 and 7 ceased to be the focus of renewable energy policy. This shift indicates a gradual transition in the focus of renewable energy policy towards national energy planning and market regulation. Governments and policymakers have recognized the importance of comprehensive energy planning and market regulation in promoting renewable energy. Governments are promoting national energy planning to integrate renewable energy with conventional energy sources to create a more sustainable energy system and ensure the healthy development of renewable energy markets.

In conclusion, the analysis conducted in this study, utilising linear segmentation and a heat map, effectively unveils the focal areas and critical issues of renewable energy policies at different stages. The findings can serve as a reference basis for policy formulation and enhancement.

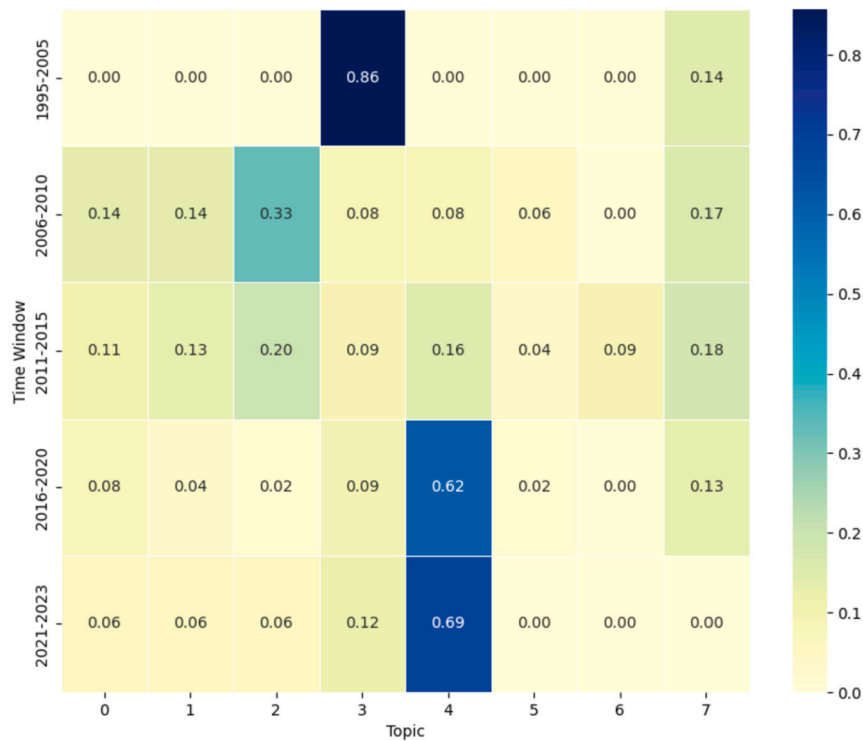


Fig. 4. Topic Intensities by Time Window.

## 5. Conclusions and policy implications

This study employs LDA topic modelling to examine the evolution and development of renewable energy policies. Initially, high-frequency keyword analysis is conducted to identify the key research focus in renewable energy policies. The analysis reveals that “development” has been the central focus of energy policy, driving the rapid growth of related industries. The insights from the keyword analysis offer valuable information for policymakers and researchers to understand the core content and areas of emphasis in renewable energy policies. Subsequently, a keyword co-occurrence network analysis is performed to uncover the interrelationships and patterns among the identified keywords. This analysis constructs a network based on the relationship between keywords, providing an intuitive depiction of the centrality and relevance of the keywords. The intense focus on renewable energy policy and the “development” keyword is a successful model for the renewable energy industry. Furthermore, the significance of “new energy” as a prominent keyword in renewable energy policies emphasizes the policy document’s attention to new energy. It reinforces the findings from the analysis of the development stage of renewable energy. The outcomes of the high-frequency keyword analysis and keyword co-occurrence network analysis consolidate our comprehension of the characteristics and focal points within the policy documents. This understanding is fundamental for policymakers to develop and enhance policies and guide decision-making processes in renewable energy. Next, LDA topic modelling analyses a corpus of renewable energy policy texts, identifying several key topics and topics about renewable energy policies. Finally, the evolution and development of renewable energy policies are elucidated through time linear segmentation and heat maps. The findings demonstrate that as renewable energy has progressed, governments and policymakers have gained a deeper understanding of the importance of comprehensive energy planning, energy transition, and electricity market regulation. This understanding has propelled their commitment to strengthening policies and regulatory measures to promote sustainable development and utilization of renewable energy. In conclusion, adopting the LDA thematic model proves to be an effective tool for studying the evolution of renewable energy policy in China. The results of this model provide valuable guidance for policymakers and stakeholders in the industry.

### CRedit authorship contribution statement

**Wenyu Ma:** Resources, Project administration, Methodology, Formal analysis. **Wenyu Wang:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



## Data availability

The data that support the findings of this study are openly available in the Peking University's Fabao Policy Database at <https://www.pkulaw.com> with the permission of the Peking University's Fabao Policy Database.

## Acknowledgements

I extend my profound gratitude to all individuals who have provided unwavering support and invaluable guidance throughout the entirety of this research undertaking. Additionally, I wish to render homage to the predecessors referenced in this work, whose conscientious endeavours have significantly contributed to the advancement of research in this domain.

## Appendix A. My appendix

**Table A.1**

Distribution of topics and characteristic words in renewable energy policies.

Topic type	Topic content	Characteristic words	Weight
Topic 0	Specialized Funding and Policy Implementation	Earmarked Funds	0.128
		Experts	0.083
		Research	0.078
		Oversight	0.072
		Approach	0.069
Topic 1	Research and application of ocean energy technologies	Institutions	0.069
		Ocean Energy	0.160
		Technology	0.141
		Declaration	0.101
		Engineering	0.083
Topic 2	Renewable energy policies in the building sector	Research	0.075
		Systems	0.059
		Architecture	0.299
		Technology	0.124
		Housing	0.099
Topic 3	Planning and utilization of renewable energy technologies	Solar	0.083
		Area	0.060
		Urban and rural	0.058
		Technology	0.145
		Planning	0.109
Topic 4	National energy planning and electricity market regulation	Solar Energy	0.105
		Utilization	0.093
		Resources	0.079
		Industries	0.064
		National Energy	0.151
Topic 5	Wind power and new energy generation projects for renewable energy companies	Electricity	0.111
		Responsibility	0.079
		Annual	0.070
		Information	0.067
		Transactions	0.063
Topic 6	Wind power and biomass power generation engineering and technology	Ltd.	0.288
		Wind Farms	0.150
		Wind Power	0.147
		New Energy	0.078
		Huaneng(Chinese company)	0.074
Topic 7	Renewable energy subsidy policies and tariff surveys	Datang (Chinese company)	0.060
		Wind Power	0.317
		Wind farms	0.154
		Limited Company	0.122
		Biomass Power Generation	0.110
		Project	0.098
		Datang (Chinese company)	0.033
		Subsidies	0.250
		Electricity Prices	0.196
		Survey	0.073
		Subsidies	0.059
		Government	0.051
		Price	0.043

## References

- [1] J. Symons, Introduction: challenges to energy security in the era of climate change, in: L. Anceschi, J. Symons (Eds.), *Energy Security in the Era of Climate Change*, Palgrave Macmillan UK, London, 2012, pp. 1–10, [http://link.springer.com/10.1057/9780230355361\\_1](http://link.springer.com/10.1057/9780230355361_1).
- [2] S. Singh, J. Ru, Accessibility, affordability, and efficiency of clean energy: a review and research agenda, *Environ. Sci. Pollut. Res.* 29 (2022) 18333–18347.
- [3] K.-m. Zhang, Z.-g. Wen, Review and challenges of policies of environmental protection and sustainable development in China, *J. Environ. Manag.* 88 (2008) 1249–1261.
- [4] D. Song, B. Jia, H. Jiao, Review of renewable energy subsidy system in China, *Energies* 15 (2022) 7429.
- [5] W. Wei, L. He, X. Li, Q. Cui, H. Chen, The effectiveness and trade-offs of renewable energy policies in achieving the dual decarbonization goals in China: a dynamic computable general equilibrium analysis, *Int. J. Environ. Res. Public Health* 19 (2022) 6386.
- [6] D. Song, Y. Liu, T. Qin, H. Gu, Y. Cao, H. Shi, Overview of the policy instruments for renewable energy development in China, *Energies* 15 (2022) 6513.
- [7] L. Li, A. Taelhagh, An in-depth analysis of the evolution of the policy mix for the sustainable energy transition in China from 1981 to 2020, *Appl. Energy* 263 (2020) 114611.
- [8] Y. Yang, K. Lo, China's renewable energy and energy efficiency policies toward carbon neutrality: a systematic cross-sectoral review, *Energy Environ.* (2023), 0958305X2311674.
- [9] D. Yu, B. Xiang, Discovering topics and trends in the field of artificial intelligence: using LDA topic modeling, *Expert Syst. Appl.* 225 (2023) 120114.
- [10] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, L. Zhao, Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey, *Multimed. Tools Appl.* 78 (2019) 15169–15211.
- [11] C. Schwarz, Ldagibbs: a command for topic modeling in Stata using latent Dirichlet allocation, *Stata J., Promot. Commun. Stat. Stata* 18 (2018) 101–117.
- [12] D. Yu, A. Fang, Z. Xu, Topic research in fuzzy domain: based on LDA topic modelling, *Inf. Sci.* 648 (2023) 119600.
- [13] Y.W. Teh, D. Newman, M. Welling, A Collapsed Variational Bayesian Inference Algorithm for Latent Dirichlet Allocation, Technical Report, Defense Technical Information Center, Fort Belvoir, VA, 2007, <http://www.dtic.mil/docs/citations/ADA629956>.
- [14] D. Andrzejewski, X. Zhu, Latent Dirichlet allocation with topic-in-set knowledge, in: *Proceedings of the NAACL HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing - SemiSupLearn '09*, Association for Computational Linguistics, Boulder, Colorado, 2009, pp. 43–48, <http://portal.acm.org/citation.cfm?doid=1621829.1621835>.
- [15] L. AlSumait, D. Barbará, J. Gentle, C. Domeniconi, Topic significance ranking of LDA generative models, in: W. Buntine, M. Grobelnik, D. Mladenić, J. Shawe-Taylor (Eds.), *Machine Learning and Knowledge Discovery in Databases*, in: *Lecture Notes in Computer Science*, vol. 5781, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 67–82, [http://link.springer.com/10.1007/978-3-642-04180-8\\_22](http://link.springer.com/10.1007/978-3-642-04180-8_22).
- [16] R. Albalawi, T.H. Yeap, M. Benyoucef, Using topic modeling methods for short-text data: a comparative analysis, *Front. Artif. Intell.* 3 (2020) 42.
- [17] L. Hagen, Content analysis of e-petitions with topic modeling: how to train and evaluate LDA models?, *Inf. Process. Manag.* 54 (2018) 1292–1307.
- [18] D. Yokomoto, K. Makita, H. Suzuki, D. Koike, T. Utsuro, Y. Kawada, T. Fukuhara, LDA-based topic modeling in labeling blog posts with Wikipedia entries, in: D. Hutchison, T. Kanade, J. Kittler, J.M. Kleinberg, F. Mattern, J.C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M.Y. Vardi, G. Weikum, H. Wang, L. Zou, G. Huang, J. He, C. Pang, H.L. Zhang, D. Zhao, Z. Yi (Eds.), *Web Technologies and Applications*, in: *Lecture Notes in Computer Science*, vol. 7234, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 114–124, [http://link.springer.com/10.1007/978-3-642-29426-6\\_15](http://link.springer.com/10.1007/978-3-642-29426-6_15).
- [19] A.P. Shiryaev, A.V. Dorofeev, A.R. Fedorov, L.G. Gagarina, V.V. Zaycev, LDA models for finding trends in technical knowledge domain, in: *2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EConRus)*, IEEE, St. Petersburg, 2017, pp. 551–554, <https://ieeexplore.ieee.org/document/7910614/>.
- [20] F. Yue, K. Guo, M. Yue, W. Chen, Topic mining and effectiveness evaluation of China's coal-related policy based on LDA model, in: Q. Liang, W. Wang, J. Mu, X. Liu, Z. Na, X. Cai (Eds.), *Artificial Intelligence in China*, in: *Lecture Notes in Electrical Engineering*, vol. 653, Springer Singapore, Singapore, 2021, pp. 573–581, [https://link.springer.com/10.1007/978-981-15-8599-9\\_66](https://link.springer.com/10.1007/978-981-15-8599-9_66).
- [21] J. Zhang, W. Gui, J. Wen, China's policy similarity evaluation using LDA model: an experimental analysis in Hebei province, *J. Inf. Sci.* (2022) 016555152210978.
- [22] D. Yu, Z. Yan, Combining machine learning and main path analysis to identify research front: from the perspective of science-technology linkage, *Scientometrics* 127 (2022) 4251–4274.