

Article

Air Pollution Regulation for Sustainable Development in China: A BERTopic Analysis of 12,081 Policies

Yifen Xia ^{1,*}  and Yuanzhuo Wu ^{1,2,*}¹ School of Public Policy and Administration, Nanchang University, Nanchang 330031, China² Jiangxi Regional Economic Research Institute, Nanchang 330031, China

* Correspondence: 352700230001@email.ncu.edu.cn (Y.X.); wuyuanzhuo@ncu.edu.cn (Y.W.)

Abstract

Air pollution governance is crucial to China's sustainable development, and the complexity of the low-carbon transition necessitates a systematic assessment of the evolving policy landscape. This study constructs a large-scale corpus of air pollution environmental regulation documents issued by the central government and 31 provincial-level governments, collected through large-scale web crawling from official government portals. This study utilizes the BERTopic model to extract policy topics and systematically analyze policies through topic content, topic hierarchy, topic similarity, and clustering structure, as well as dynamic topic evolution over time. Key findings are as follows: (1) The BERTopic results identify 31 policy topics, and Topic 0–Topic 11 collectively account for 82.8% of all documents. (2) The topics mainly focus on three areas: comprehensive air pollution control and enforcement, structural low-carbon and energy transition, and governance capacity building through monitoring, fiscal incentives, and carbon accounting. (3) Hierarchical and similarity analyses indicate three relatively stable core thematic groups, alongside two specialized peripheral themes. (4) Over the study period, dynamic topic trends show a shift in policy emphasis from air quality-oriented control toward low-carbon transition and supporting policy instruments. These findings clarify the thematic structure and evolution of China's environmental policies, offering an evidence base for improving integrated air pollution governance toward sustainable development.

Keywords: air pollution control; policy text mining; environmental regulation policy; BERTopic model; China

1. Introduction

During China's rapid industrialization and urbanization, remarkable economic growth has been accompanied by long-standing air pollution challenges, with fine particulate matter, especially PM_{2.5}, being a prominent concern [1]. Air pollution not only threatens public health and leads to premature mortality, but can also hinder high-quality development through channels such as reduced labor productivity, increased medical expenditures, and altered location and migration choices [2]. Therefore, achieving sustained improvement under ongoing economic restructuring and energy system transition remains a core task within China's ecological civilization agenda [3].

In response to persistent air pollution, China has enacted a series of strict regulatory policies. The 2013 Air Pollution Prevention and Control Action Plan marked a new governance phase by setting binding targets and spurring swift implementation across



Academic Editor: Pallav Purohit

Received: 27 January 2026

Revised: 20 February 2026

Accepted: 22 February 2026

Published: 26 February 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

regions [4]. Building on this, the 2014 revision of the Environmental Protection Law significantly increased penalties for illegal pollutant discharge [5]. The Blue Sky Defense Battle, implemented from 2018 to 2020, combined regional joint prevention and control with residential coal substitution and tighter controls on mobile sources, contributing to a further step change in air quality outcomes [6]. Since the onset of the Fourteenth Five-Year Plan, the policy orientation has shifted from end-of-pipe treatment to source-based and structural emission reductions. It promotes energy structure optimization, green industrial transformation, and the digitalization of governance methods [7]. Recent observation-based assessments have shown that these sustained efforts have delivered substantial nationwide air quality improvements over the past decade [8], including pronounced declines in fine-particulate pollution and fewer severe pollution events [9].

Existing studies have conducted extensive empirical evaluations of air pollution governance in China from the perspectives of policy instruments, enforcement and supervision, and environmental performance, generating a wealth of evidence on determinants and mechanisms [10]. First, many studies treat policy as an exogenous shock or an explanatory variable, focusing on one policy program or a limited set of policy samples to identify average effects. However, the thematic structure, agenda evolution, and underlying governance logic embedded in the policy texts themselves have not been systematically characterized [11,12]. Second, at the methodological level, traditional topic models such as Latent Dirichlet Allocation (LDA) have limited capacity to capture word order and contextual semantics, which can lead to blurred topic boundaries and unstable representations [13]. To address these gaps, this study applies the BERTopic model to identify topics and analyze topic evolution in 12,081 air pollution environmental regulation policy texts issued between 2006 and 2024 by China's central government and 31 provincial-level governments, providing evidence to support policy optimization for long-term governance goals.

The contributions of this study can be summarized in three respects. First, BERTopic is applied to air pollution environmental regulation policies using the qwen3-embedding model to represent long documents. This improves semantic representation beyond bag-of-words features and alleviates the short text limitation common in prior BERTopic applications, enhancing topic coherence and interpretability. Second, a comprehensive policy-text database is constructed covering the central government and 31 provincial-level administrative units. Compared with studies relying on partial coverage or single-level samples, this dataset enables systematic comparisons across regions and governance levels and supports a more complete mapping of the national policy topic structure. Third, policy implications are derived from the identified thematic evolution, providing evidence-based guidance for advancing long-term sustainable environmental governance.

The remainder of this study is organized as follows. Section 2 reviews the literature on air pollution, and advances in text analysis methods. Section 3 describes the data sources, preprocessing procedures, and the design of the BERTopic model. Section 4 reports the topic content and hierarchical structure, topic clustering patterns, and dynamic topic evolution over time. Section 5 concludes with the main findings and policy implications.

2. Literature Review

2.1. Research on Environmental Regulation Policies

Environmental regulation generally refers to a set of laws, policies, and administrative measures adopted by governments to control pollution and protect the environment [14]. In practice, environmental regulation is commonly categorized into three types: command-and-control regulation [15], market-based regulation [16], and participatory regulation [17]. Command-and-control regulation relies on mandatory rules and direct enforcement, such as emission standards, total emission control, and environmental permitting [18]. Under

this framework, authorities specify environmental quality targets or emission limits and require strict compliance through routine inspections and administrative penalties [19]. Market-based regulation applies economic incentives, including pollution charges, environmental taxes, emissions trading, and subsidies or rewards, to encourage firms to reduce emissions in a cost-effective manner [20,21]. Voluntary or participatory measures involve corporate environmental responsibility, environmental information disclosure, and public participation [22], whereby reputational pressures and social oversight motivate polluters to take actions beyond minimum legal requirements [23].

Regarding the effectiveness of environmental regulation, the empirical findings remain mixed. Some studies support the Porter hypothesis [24], suggesting that appropriately stringent environmental regulation can stimulate firms to innovate and improve resource-use efficiency [25]. Such innovation offsets the costs of compliance, thereby achieving a “win-win” outcome of environmental improvement and economic growth [26]. However, other research argues that overly stringent environmental requirements may increase compliance costs and weaken economic efficiency, potentially inducing a “green paradox” in specific contexts [27]. Moreover, environmental regulation often exhibits a nonlinear impact on performance, maintaining an inverted U-shaped relationship with green innovation, while following a U-shaped trajectory with carbon intensity [28].

In summary, existing research indicates that environmental regulation policies have played a generally positive role in pollution control, but their effectiveness depends on policy design and the strength of implementation and enforcement [29]. Key pathways for improving regulatory performance include setting scientifically grounded environmental standards, strengthening the legal framework and market-based mechanisms, enhancing enforcement supervision and transparency, and promoting cross-regional cooperation for joint prevention and control [30]. Focusing on the air pollution domain in China, this study applies topic-based text analysis to map policy emphases and their evolutionary patterns across different stages, thereby responding to the limited evidence on systematic policy agenda evolution.

2.2. Research on Policy Text Analysis

Policy texts contain rich information, and the systematic analysis of policy documents helps reveal patterns of policy evolution and shifts in governmental priorities [31]. However, conventional policy analysis has largely relied on qualitative reading or manual classification, which is difficult to scale to large corpora and is susceptible to researcher subjectivity and coding bias [32]. With recent advances in natural language processing, an increasing number of studies have introduced text-mining approaches into policy research to enable more objective and reproducible quantitative analyses of policy documents [33].

Topic modeling is among the most widely used text-mining tools for automatically extracting latent themes from large document collections. Latent Dirichlet Allocation (LDA) is a classic topic modeling method that represents each document as a probabilistic mixture of topics, and each topic as a probability distribution over words, thereby uncovering the topic structure without manual labeling [34]. Over the past two decades, LDA has inspired numerous extensions and has been widely applied to policy, management, and other social science text analyses [35]. Some scholars have combined LDA-based topic extraction with time-series forecasting methods to predict future trends in topic distributions within policy texts [36].

However, traditional topic models such as Latent Dirichlet Allocation are built on a bag-of-words assumption and do not model word order or contextual semantics, which limits their ability to represent meaning-rich policy documents and may lead to less informative topic representations [37]. With the rapid development of deep learning in recent years,

topic analysis methods based on pretrained language models have emerged. Among them, BERTopic frames topic modeling as a semantic clustering process [38]. Empirical comparisons in applied settings further suggest that BERTopic can yield topic clusters with improved separability and higher semantic coherence than LDA, benefiting from contextual embeddings [39]. In addition, BERTopic supports extensions such as topic evolution analysis over time and hierarchical topic organization, enabling the investigation of temporal shifts and hierarchical relationships among topics [40].

Beyond topic modeling, researchers have applied a range of text-analytic techniques to policy studies [41]. For example, Structural Topic Models (STM) enable the incorporation of document metadata as covariates, thereby facilitating systematic comparisons of topic prevalence and topical content across policy attributes [42]. In research on policy evolution, topic model outputs have also been combined with time-series approaches [43]. In addition, frequency-based methods and co-word or semantic network analyses have been widely used to identify salient policy terms and to reveal association structures among policy concepts [44].

In conclusion, these approaches provide quantitative lenses for interpreting policy texts and for reducing subjectivity in qualitative coding. Accordingly, this study employs BERTopic to analyze China's air pollution environmental regulation policy corpus and to produce more objective and reproducible evidence. A key adaptation is the use of semantic embeddings generated by a Qwen3-Embedding model, which can represent complex and lengthy policy documents with richer contextual information. This helps address a common limitation in earlier BERTopic applications that primarily focused on short texts such as online comments, tweets, or brief news snippets.

3. Research Design

3.1. Research Framework

This study examines the thematic structure and temporal evolution of China's air pollution environmental regulation policies using large-scale policy text mining. The research process follows four steps. First, policy documents issued by the central government and 31 provincial-level governments are mainly obtained by crawling official government websites, and they are then supplemented with documents from authoritative databases. Second, the corpus is preprocessed through text cleaning, word segmentation, stop-word removal, and the standardization of metadata such as issuing body, region, and release date. Third, the BERTopic model is trained to generate policy topics, and topic interpretation is refined by integrating keyword weights and representative documents to enhance semantic coherence and readability. Fourth, topic evolution is assessed by tracing annual topic proportions and document counts, and topic relationships are further examined through hierarchical and similarity analyses to identify shifts in policy priorities over time. The research framework is shown in Figure 1.

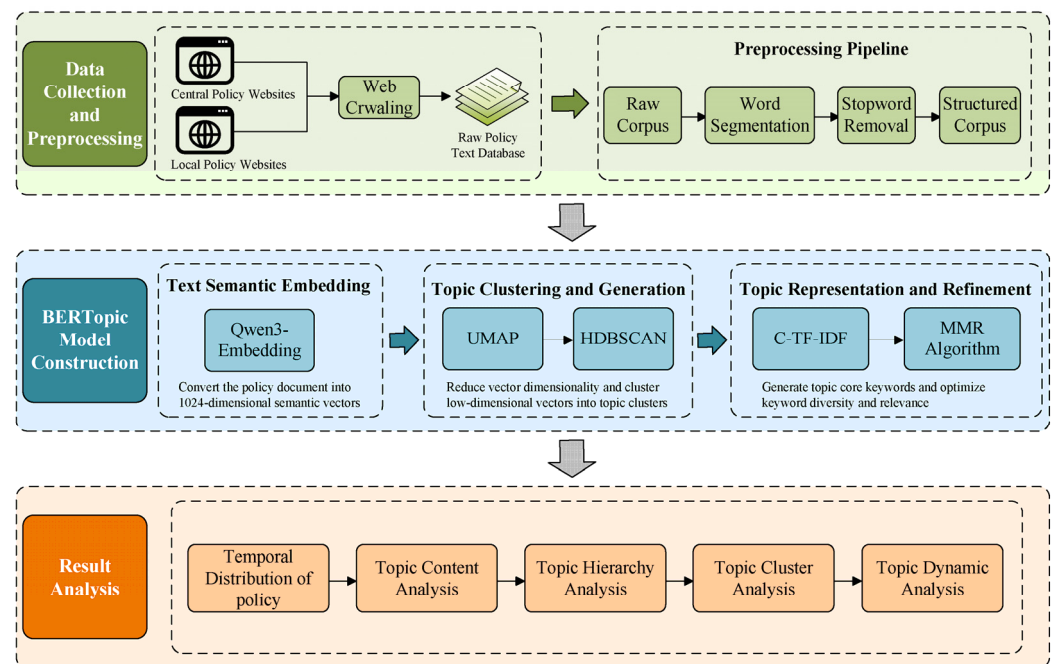


Figure 1. Research framework for topic modeling and evolution analysis.

3.2. Data

The policy text dataset used in this study consists of environmental regulation policy documents issued by China's central and local governments between 2006 and 2024. Central-level documents mainly include laws and administrative regulations, State Council documents, ministerial rules, and other normative policy instruments issued by the State Council and its affiliated ministries and commissions. Local level documents cover 31 provincial-level administrative units, including provinces, autonomous regions, and municipalities, and comprise local regulations, local government rules, and local policy and implementation documents. Data were collected using a dedicated web crawler to support the large-scale acquisition of policy texts from official government portals. We built automated crawlers to systematically collect environmental regulation policies issued by China's central government and 31 provincial-level governments from 2006 to 2024. Because website structures and publication formats vary across jurisdictions, we used a keyword-based retrieval strategy with environment-related terms, including environmental protection, pollution prevention and control, energy saving and emission reduction, ecological civilization, and low carbon. We also implemented adaptive extraction rules to capture full texts and to standardize key metadata, such as the issuing authority, document number, and release date. Next, we further removed duplicate entries caused by cross-site reposting so that each record represented a unique policy document. The final dataset consists of 12,081 policy documents. As shown in Table 1, there are 491 documents at the central level, while the 11,590 local-level documents are detailed in Table 2.

Following data collection, the texts were preprocessed using a standard pipeline. All documents were converted into plain text, and non-body content such as HTML tags, annex tables, and icons was removed. The Chinese texts were then tokenized using the Jieba tokenizer to segment continuous character sequences into word tokens. To improve tokenization quality and topic interpretability, a domain-specific stop-word list and a user-defined dictionary were constructed, removing common stop words and retaining domain-specific environmental terms. Finally, selected proper nouns were merged using synonym normalization to reduce fragmentation. After preprocessing, a high-quality

policy corpus was established, comprising approximately 85.52 million Chinese characters, providing a solid foundation for subsequent topic modeling.

Table 1. China’s air pollution environmental regulation policies at the central level (excerpts).

Year	Policy Name	Policy Number	Policy Issuer
2006	Notice on Issuing the Interim Measures for Public Participation in Environmental Impact Assessment	Huanfa [2006] No. 28	State Environmental Protection Administration
⋮	⋮	⋮	⋮
2024	Notice of the General Office of the State Council on Issuing the Work Plan for Accelerating the Establishment of the Dual-Control System for Carbon Emissions	Guobanfa [2024] No. 39	General Office of the State Council

Table 2. China’s air pollution environmental regulation policies at the local level (Beijing).

Year	Policy Name	Policy Number	Policy Issuer
2006	Notice of the General Office of the Beijing Municipal People’s Government on Assigning Tasks for Implementing the 12th-Stage Measures for Air Pollution Control	Jingzhengbanfa [2006] No. 13	Beijing Municipal People’s Government
⋮	⋮	⋮	⋮
2024	Notice of the Beijing Municipal People’s Government on Issuing the Measures for the Administration of Beijing Carbon Emissions Trading	Jingzhengfa [2024] No. 6	Beijing Municipal People’s Government

3.3. Method

To extract interpretable policy topics from a large-scale corpus and to analyze their structure and evolution, a BERTopic modeling pipeline was established. For each topic, the model outputs representative keywords with corresponding weights, and it also assigns each document a topic label and an associated probability, which are used for subsequent analyses. The BERTopic pipeline in this study consists of five components: semantic embedding, UMAP-based dimensionality reduction, HDBSCAN clustering, c-TF-IDF topic representation, and MMR-based keyword diversification.

First, a pretrained embedding model was used to map each policy text into a high-dimensional continuous vector space. Specifically, the Qwen3-Embedding model was adopted, which is trained on large-scale multilingual corpora and is designed to capture semantic similarity, such that sentences with similar meanings are mapped to nearby locations in the embedding space. Given a policy document set D , each document $d_i \in D$ is encoded into an embedding vector e_i :

$$e_i = f_{Qwen}(d_i), e_i \in \mathbb{R}^{1024} \quad (1)$$

This step projects semantically similar policy expressions into neighboring regions of the vector space and provides informative features for clustering.

Since the embedding vectors have a dimensionality of 1024, direct clustering in the original space may be hindered by the curse of dimensionality. Therefore, Uniform Manifold Approximation and Projection (UMAP) was employed to perform nonlinear dimensionality reduction. UMAP assumes that the data are approximately uniformly distributed

on a manifold, and it seeks a low-dimensional representation by minimizing the fuzzy cross-entropy between the high-dimensional and low-dimensional topological structures:

$$C = \sum_{i \neq j} \left(p_{ij} \log \frac{p_{ij}}{q_{ij}} + (1 - p_{ij}) \log \frac{1 - p_{ij}}{1 - q_{ij}} \right) \quad (2)$$

where p_{ij} and q_{ij} denote the pairwise similarities between samples in the high-dimensional and low-dimensional spaces, respectively. In this study, the parameters $n_neighbors$ and min_dist were tuned to balance the preservation of local neighborhoods and the global topological structure in the policy text embedding space.

Topic clustering was then performed using HDBSCAN in the reduced embedding space. HDBSCAN identifies clusters of arbitrary shapes while filtering noise by constructing a density-based minimum spanning tree. In HDBSCAN, documents that fall into low-density regions are labeled as outliers (Topic -1). In this study, these outliers were treated as noise points and were not reassigned to any smaller topic; therefore, they were excluded from the statistics of Topic 0–Topic 30. Its core procedure relies on computing the mutual reachability distance between samples:

$$d_{mreach,k}(a, b) = \max\{core_k(a), core_k(b), d(a, b)\} \quad (3)$$

where $core_k$ denotes the core distance. In this study, $min_cluster_size$ was set to 60 to balance thematic granularity with noise reduction. This moderate threshold helps mitigate the overshadowing effect, ensuring that smaller but significant regulatory topics are preserved rather than being absorbed into broader, high-frequency categories. This step groups semantically dense policy texts into K preliminary topic clusters.

To extract discriminative keywords from each topic cluster, class-based TF-IDF (c-TF-IDF) was applied. This method treats all documents within the same topic cluster as an aggregated class and computes the importance of a term x in class c :

$$W_{x,c} = TF_{x,c} \times \log \left(1 + \frac{A}{f_x} \right) \quad (4)$$

where $TF_{x,c}$ denotes the frequency of term x in cluster, f_x represents the total frequency of the term across the entire corpus, and A is the average number of words per corpus. The c-TF-IDF effectively reduces the weights of common administrative terms and highlights technical and policy-relevant terminology that is distinctive for each topic.

Although c-TF-IDF can identify terms that are highly relevant to a topic, the top-ranked keywords may exhibit semantic redundancy, which can limit the diversity of topic descriptions. Maximal Marginal Relevance (MMR) was adopted to re-rank candidate keywords and promote diversification. By penalizing redundant terms, MMR enhances the distinctiveness of the model. This ensures a balanced and comprehensive interpretation of the entire policy landscape, allowing even smaller topics to maintain clear and unique thematic profiles:

$$MMR = \arg \max_{D_i \in R \setminus S} \left[\lambda \cdot \text{Sim}_1(D_i, Q) - (1 - \lambda) \cdot \max_{D_j \in S} \text{Sim}_2(D_i, D_j) \right] \quad (5)$$

where Q represents the topic vector, D_i denotes the candidate terms, S is the set of previously selected keywords, and λ acts as the diversity factor. The λ parameter allows for the adjustment of keyword heterogeneity. With λ set at 0.3, this study filters out semantically repetitive terms while maintaining high alignment with policy themes. This approach ensures that the 10 keywords describe environmental policy themes from diverse perspectives, thereby improving the model's semantic coverage and interpretability.

In addition, a human validation step was conducted to support the interpretability of the extracted high-frequency topics. Specifically, 30 policy documents were randomly sampled from each high-frequency topic, and their substantive consistency with the assigned topic name was checked through close reading. This procedure was used to verify whether the topic naming was coherent with the underlying documents and to reduce the risk of mislabeling during topic interpretation.

To ensure the transparency and reproducibility of this study, the analytical framework and parameter configurations for the BERTopic model, including embedding, dimensionality reduction, clustering, and representation, are fully disclosed. The specific settings for each stage are detailed in Table 3.

Table 3. Model Parameters of the BERTopic.

Step	Parameter	Value
Embedding	SentenceTransformer	Qwen3-Embedding
DimensionalityUMAP	n_neighbors	20
	n_components	5
	metric	cosine
ClusteringHDBSCAN	min_cluster_size	60
	min_samples	3
	metric	Euclidean
CountVectorizer	min_df	20
	ngram_range	(1, 1)
MMR	diversity	0.3

4. Results and Discussion

4.1. Temporal Distribution of Policy

Figure 2 shows the annual number of air pollution environmental regulation policies issued by China's central government from 2006 to 2024, with shaded backgrounds indicating the Eleventh to Fourteenth Five-Year Plan periods. Central policy output is relatively small and follows a fluctuating upward trend. During the Eleventh Five-Year Plan, policy numbers remained low, suggesting air pollution control was not yet a nationwide priority. In 2008 and 2009, amid economic downturn, the central government focused more on economic growth and industrial restructuring.

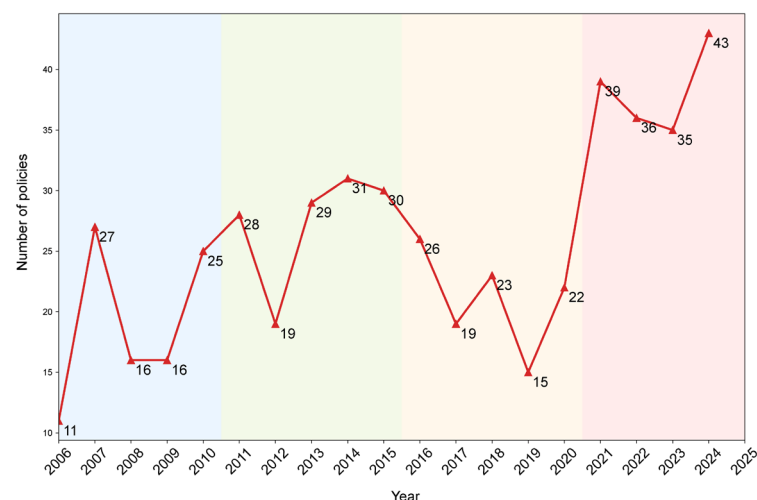


Figure 2. The trend of the timing of central policy announcements.

In the Twelfth Five-Year Plan period, central policy issuance increased steadily, with around 30 policies per year in most years except 2012. This phase was guided by the Action

Plan for Air Pollution Prevention and Control released by the State Council in September 2013. During the Thirteenth Five-Year Plan period, annual issuance stayed relatively stable in the range of 15 to 25 policies. The Three-Year Action Plan for Winning the Blue Sky Defense War, issued by the State Council in 2018, coincided with an increase of four policies compared with 2017.

After China announced its carbon peaking and neutrality goals in 2020, policy priorities refocused on air environment governance. During the Fourteenth Five-Year Plan, central policy issuance increased, reaching its peak in 2024. The current agenda continues improving air quality while adding requirements for coordinated pollution reduction and carbon mitigation.

Figure 3 shows the annual total of air pollution regulations by 31 provincial governments. Local policy output greatly exceeds the central government and shows a stronger growth trend. During the Eleventh Five-Year Plan, local issuance was low and increased gradually, mainly targeting specific industries or pollution sources without an institutionalized framework.

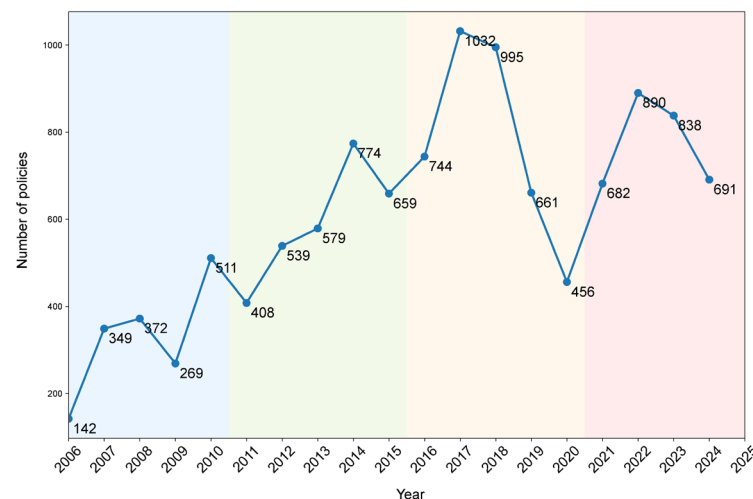


Figure 3. The trend of the timing of local policy announcements.

During the Twelfth Five-Year Plan, local policy issuance rose sharply, peaking at 774 documents in 2014. After the central government issued the Action Plan for Air Pollution Prevention and Control in 2013, provinces increased supporting policies to align with central targets and aid implementation. In the Thirteenth Five-Year Plan, local issuance peaked in 2017, stayed high in 2018, but declined clearly in 2019. This trend matches the policy cycle of the Three-Year Action Plan for Winning the Blue Sky Defense War, launched in 2018. It signals a shift from producing new policies to routine implementation after specific tasks are set.

The three-year decline reached a turning point in 2020, and policy issuance resumed a fluctuating upward trend during the Fourteenth Five-Year Plan period. After the Action Plan for Continuous Air Quality Improvement was released in 2023, provincial governments needed to develop province specific implementation plans and integrate pollution reduction and carbon mitigation with energy structure adjustment within a unified governance framework.

4.2. Topic Content Analysis

This study employs BERTopic to extract topics from air pollution environmental regulation policy texts issued by China's central government and 31 provincial-level governments. A total of 31 distinct topics were identified, accounting for 96.24% of the analyzed policy corpus. Because large differences in topic size may affect the stability of interpretation,

the analysis focuses on high-frequency topics with more than 200 documents (Topic 0–Topic 11). These 12 topics cover 9625 policy texts, accounting for 82.8% of the full corpus, and therefore provide sufficient coverage of the main policy concerns in China’s air pollution governance. To facilitate interpretation, Topics 0–11 are named based on their most representative keywords, and the topic labels together with the top 10 keywords for each topic are reported in Table 4.

Table 4. The topic results of BERTopic and the top ten keywords.

Topic	Name	Top Words
Topic 0	Coal Dust Control	coal combustion, air quality, dust, air pollution, regulation, phase out, law enforcement, blue sky, integrated control, monitoring
Topic 1	Dual-Carbon Pathway	climate change, low carbon, green low carbon, carbon peak, carbon neutrality, innovation, greenhouse gases, carbon emissions, low carbon development, climate
Topic 2	Efficiency Upgrading	energy saving, emission reduction, energy conservation, energy consumption, energy use, energy intensity, standard coal, inspection, energy efficiency, supervision
Topic 3	Monitoring Disclosure	environmental monitoring, monitoring data, environmental quality, monitoring sites, pollution sources, monitoring technology, information disclosure, monitoring network, falsification, centralized
Topic 4	Cleaner Production	cleaner production, economic benefits, disclosure, trade and commerce, informatization, chemicals, hazardous substances, technology, performance, pollution reduction
Topic 5	Subsidy Funds	special funds, funds, rewards, subsidies, energy saving, emission reduction, transportation, science and technology, budget, investment
Topic 6	Carbon Trading MRV	verification, carbon emissions, greenhouse gases, allowances, accounting, trading, CO ₂ , emission volume, validation, trading market
Topic 7	Straw Burning Control	straw, dust, air pollution, air quality, responsibility, grid-based management, integrated control, regulation, law enforcement, oil and gas
Topic 8	Renewables Integration	new energy, photovoltaics, wind power, power grid, renewable energy, power generation, electricity, batteries, power stations, solar
Topic 9	Renewable Subsidies	renewable energy, solar, subsidies, funds, photovoltaics, special funds, fiscal, building energy efficiency, review, supervision
Topic 10	Air Enforcement Penalties	fines, atmospheric environment, supervision and management, air pollution, business suspension, fugitive dust, pollutants, total emission control, ambient standards, management
Topic 11	Complaint Sanctions	fines, environmental impact, pollution issues, reporting, law, disputes, administrative penalties, construction projects, supervision, management

4.2.1. Topic Keyword Decay

Figure 4 shows the decay trend of keyword weights as the keyword rank increases within each topic. Considering the aggregate pattern for Topic0 to Topic 11, the mean keyword weight is approximately 0.479 for rank 1, decreases to around 0.300 by rank 5, and further declines to about 0.247 by rank 10, with an obviously slower decay rate thereafter. This pattern suggests that the core semantics of a topic are primarily conveyed by a small number of high-weight keywords, whereas lower-ranked terms primarily provide supplementary information.

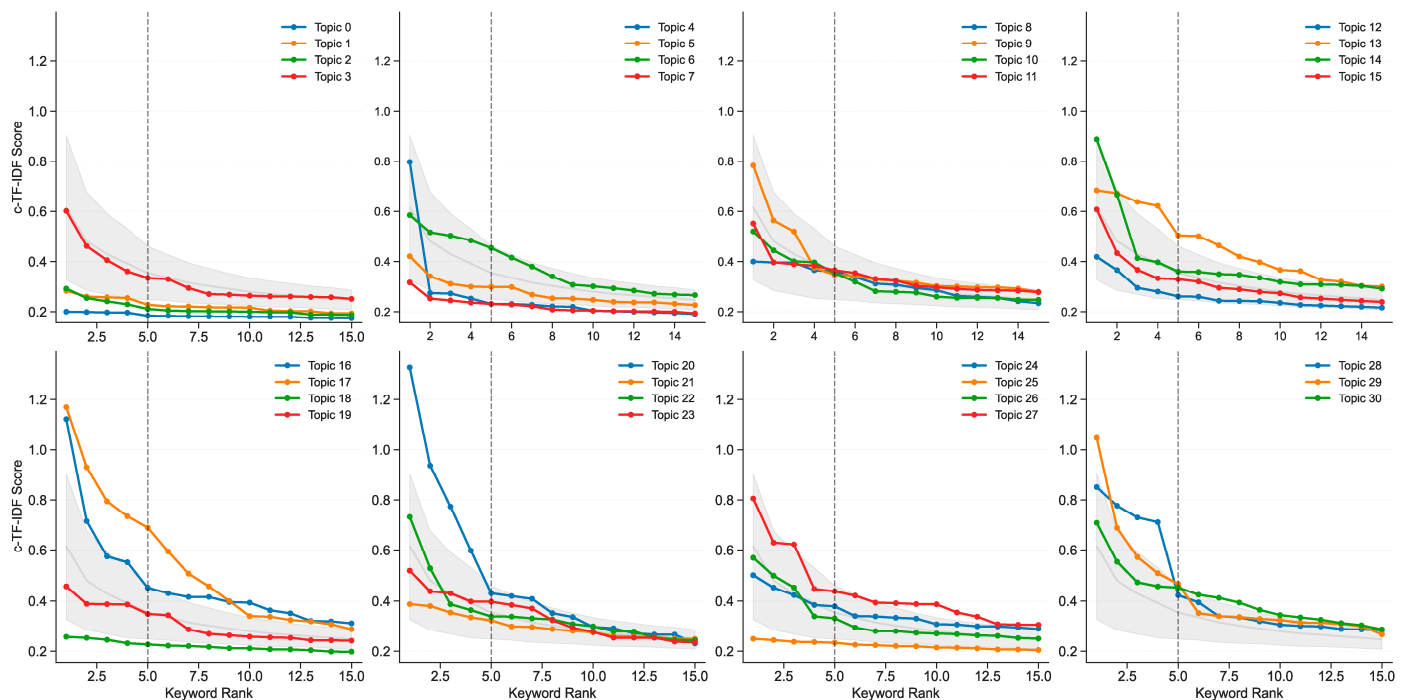


Figure 4. C-TF-IDF keyword weight decay for the topics.

Meanwhile, the decay shapes differ across topics. For example, Topic 0 and Topic 8 show relatively similar weights among the top five keywords and a more gradual decay, indicating that their topic semantics are jointly supported by multiple keywords. In contrast, Topics 4 and 9 exhibit a much higher weight for the first-ranked keyword than for subsequent keywords and thus display a steeper decay. This suggests that these topics are dominated by a single core policy anchor term, and the top five keywords are sufficient to capture the key semantic signal. Based on these patterns, this study utilizes the top five keywords for topic interpretation in the topic content analysis, which balances semantic coverage and interpretability.

4.2.2. Topic Word Scores

Figures 5–8 further report the top five keywords and their c-TF-IDF weights for Topic 0 to Topic 11. Overall, the keyword-weight structures exhibit two typical patterns across topics. In the first pattern, the weights of the top five keywords are relatively close, indicating that topic semantics are jointly supported by multiple elements and the semantic structure is more balanced; representative examples include Topic 0, Topic 2, and Topic 8. In the second pattern, a clear “dominant keyword” emerges, where the first-ranked keyword has a substantially higher weight than the remaining keywords; typical cases include environmental monitoring in Topic 3, cleaner production in Topic 4, and renewable energy in Topic 9. These two patterns correspond to different policy characteristics. The first

pattern is more consistent with policy narratives of comprehensive control and multi-source coordination, whereas the second pattern is more indicative of specialized institutional arrangements or specific industrial policy domains.

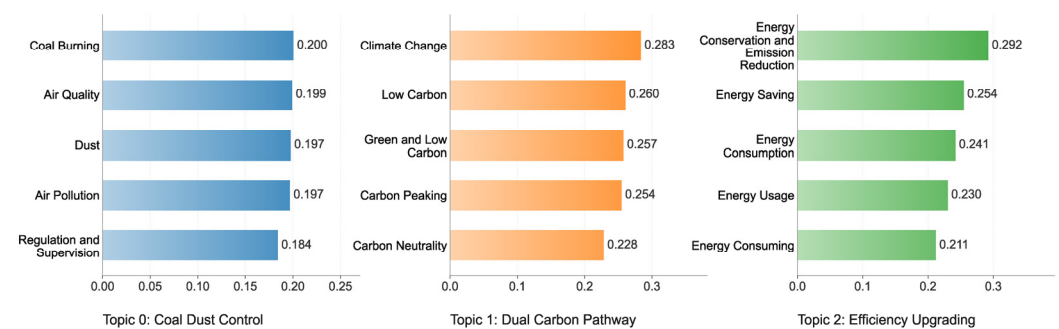


Figure 5. Top topics and subject terms of Topic 0–Topic 2.

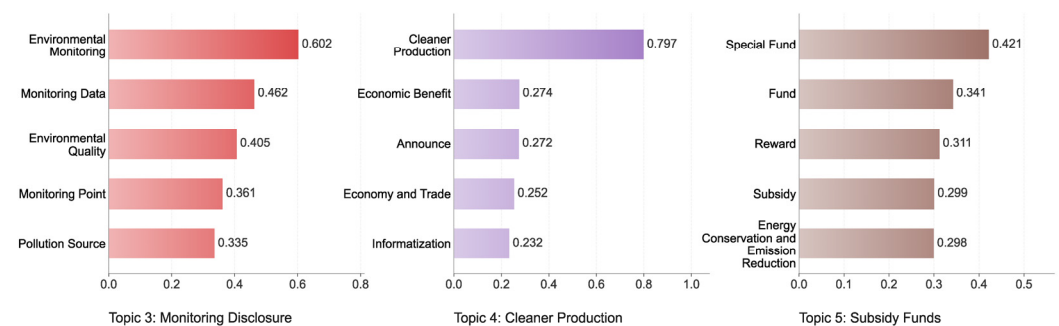


Figure 6. Top topics and subject terms of Topic 3–Topic 5.

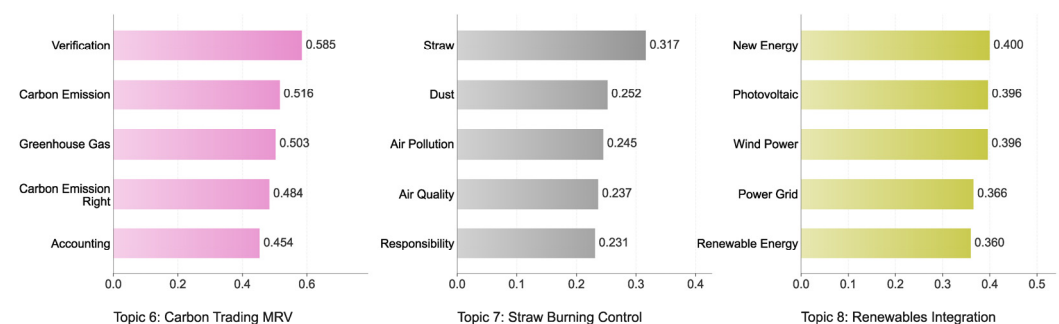


Figure 7. Top topics and subject terms of Topic 6–Topic 8.

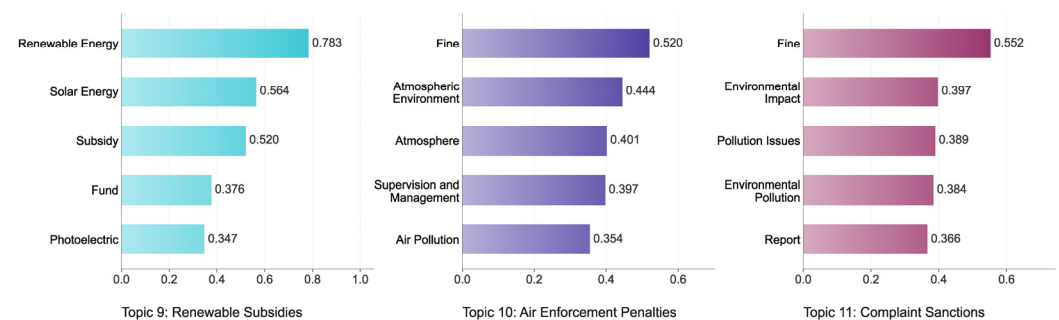


Figure 8. Top topics and subject terms of Topic 9–Topic 11.

As shown in Figure 5, Topic 0 focuses on comprehensive governance aimed at reducing coal-related emissions and improving air quality. Its top five keywords are coal combustion, air quality, dust, air pollution, and regulation, with a relatively balanced weight distribution, suggesting that policies within this topic address multiple pollution

sources while simultaneously emphasizing regulatory enforcement, reflecting an air quality-centered comprehensive governance orientation. Topic 1 focuses on climate governance and the low-carbon transition. Its keywords include climate change, low carbon, green low carbon, carbon peaking, and carbon neutrality, and the semantics are concentrated around the “dual carbon” target framework, indicating that air pollution regulation texts increasingly reinforce a long-term strategic orientation toward the synergistic control of conventional pollutants and carbon emissions. Topic 2 represents energy conservation and energy-intensity constraints. Its keywords include energy saving and emission reduction, energy conservation, energy consumption, energy use, and energy intensity, highlighting the need for efficiency improvement and energy management requirements.

As shown in Figure 6, Topic 3 keywords include environmental monitoring, monitoring data, environmental quality, monitoring sites, and pollution sources, with environmental monitoring receiving a notably higher weight. This indicates that monitoring-system development serves as a key enabling component in policy texts, providing an evidentiary basis for pollution-source management. Topic 4 focuses on cleaner production and industrial green upgrading. The keywords include cleaner production, economic benefits, disclosure, trade and commerce, and informatization. The relatively high weight of the core term cleaner production indicates that policies under this topic tend to promote process retrofitting and technological upgrading through institutionalized instruments. Topic 5 reflects fiscal funding and incentive support. Its keywords include special funds, funds, rewards, subsidies, energy saving, and emission reduction, suggesting that fiscal support and incentive-based instruments are frequently bundled with energy saving and emission reduction tasks to reduce implementation costs.

As shown in Figure 7, Topic 6 centers on carbon emission accounting, verification, and compliance governance. The keywords include verification, carbon emissions, greenhouse gases, carbon emission allowances, and accounting, indicating that the governance logic extends from target articulation toward system building, with an emphasis on monitoring, reporting, and verification, as well as allowance management, serving as institutional foundations for policy operation. Topic 7 corresponds to scenario-specific governance, such as straw burning and dust control. Its keywords include straw, dust, air pollution, air quality, and responsibility. Policies in this topic are closely aligned with specific pollution scenarios. Topic 8 focuses on renewable energy and power system support. Its keywords include renewable energy, photovoltaics, wind power, power grid, and renewable resources, indicating that policy texts address not only renewable substitution but also the development of system-carrying capacity, such as grid integration.

As shown in Figure 8, Topic 9 emphasizes the promotion and support of renewable energy. The keywords include new energy, photovoltaics, wind power, power grid, renewable energy, highlighting the role of fiscal support and suggesting that incentive-based instruments are effectively employed during the expansion of renewable energy. Topic 10 points to administrative supervision and punitive enforcement. The keywords include fines, atmospheric environment, atmosphere, supervision and management, and air pollution, underscoring hard constraint instruments centered on regulatory supervision and penalties. Topic 11 focuses on environmental impacts and public supervision. Its keywords include fines, environmental impact, environmental pollution, and reporting. By highlighting public reporting channels in addition to penalties, this topic suggests that mechanisms for social oversight and problem feedback have become widely incorporated into policy texts.

4.3. Topic Hierarchy Analysis

To reveal the latent hierarchical structure and inter-topic associations, pairwise topic distances were constructed based on topic representation vectors, and hierarchical cluster-

ing was performed. The resulting topic hierarchy is visualized as a dendrogram in Figure 9. A smaller distance indicates that two topics are closer in the semantic space, implying stronger co-occurrence patterns or shared governance contexts in policy narratives. In contrast, a larger distance suggests more pronounced differences in governance objectives, pollution targets, or policy instruments.

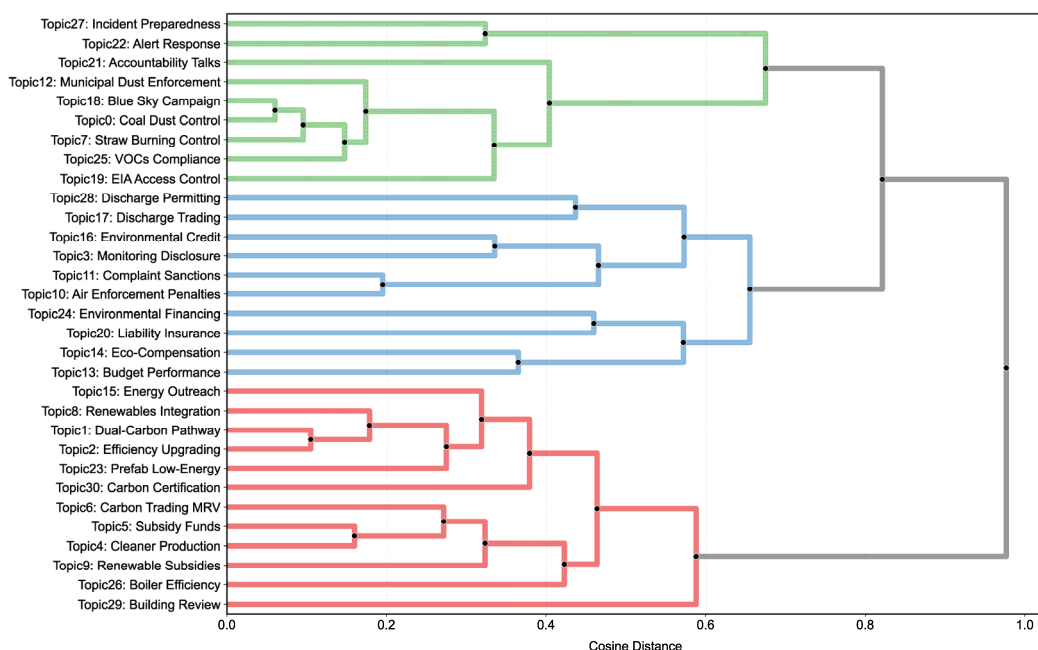


Figure 9. Hierarchical clustering structure of policy topics.

In the green merging branch of Figure 9, Topic 0 and Topic 18 merge at the smallest distance, indicating the closest semantic linkage. This tight linkage suggests that air-quality-centered comprehensive governance goals are often advanced in parallel with campaign-style actions, particularly those aligned with coal substitution and clean heating practices. This pattern is closely related to winter clean heating and scattered coal substitution policies in northern China, where key plans prioritize reducing scattered coal burning and increasing the share of clean heating, directly serving autumn–winter pollution control and air quality improvement [45]. Additionally, Topic 0 merges with Topic 7 at a relatively short distance. Topic 7 is characterized by keywords such as straw, dust, air pollution, air quality, and responsibility. Their early aggregation suggests that non-point source issues, such as open burning of straw and dust control, are frequently incorporated into the same air quality-oriented comprehensive governance framework and are linked with regulatory and enforcement instruments through responsibility assignment, inspections, and supervision.

In the blue merging branch of Figure 9, Topics 10 and Topic 11 merge at a relatively small distance, reflecting the close linkage between punitive enforcement and public supervision. Their adjacency in the hierarchy suggests that policy texts often strengthen the cost of non-compliance while simultaneously institutionalizing supervision channels, information discovery, and mechanisms for social participation, jointly supporting implementation capacity. On the one hand, improvements in the environmental rule-of-law framework have increased the costs of illegal emissions, making penalties and supervision more stable and frequent in policy texts. On the other hand, China has established and continuously improved complaint and reporting channels, forming governance procedures that cover case acceptance, transfer, and handling, and feedback on outcomes [46].

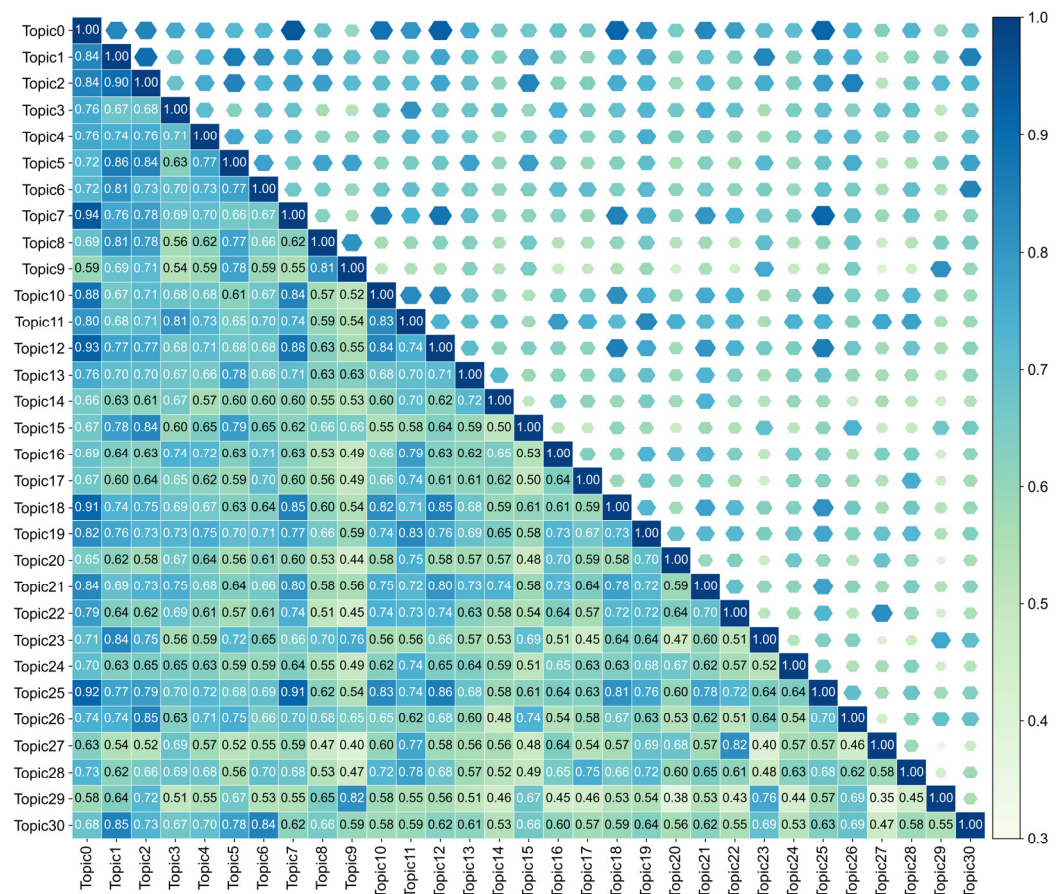
In the red merging branch of Figure 9, Topic 2 and Topic 1 merge at a small distance, indicating a strong semantic association between energy saving and energy intensity

constraints and the dual carbon goals. Their early merging in the hierarchy implies that, in policy narratives, low-carbon transition goals are often operationalized through pathways such as energy intensity constraints, energy efficiency improvements, and energy savings and emission reductions, thereby forming a stable semantic cluster. This linkage has become particularly salient following China's announcement of its carbon peaking and carbon neutrality targets, when policies have increasingly emphasized binding indicators to drive emission reductions [47].

4.4. Topic Cluster Analysis

4.4.1. Topic Similarity

To characterize the strength of semantic associations between topics, cosine similarity was computed between topic representation vectors to construct a topic similarity matrix, which is visualized as a heatmap in Figure 10. Darker colors indicate higher similarity, implying that two topics share more consistent semantic elements in policy texts, whereas lighter colors indicate larger semantic differences.



Topic 0–Topic 25 (0.917), Topic 0–Topic 18 (0.914), and Topic 0–Topic 10 (0.885). These results indicate that air quality-oriented policy narratives tend to bundle multi-source pollution control, campaign style actions, and regulatory enforcement instruments, forming a tightly connected network of related topics. In addition, low carbon transition themes also show high relatedness. For example, the similarity between Topic 1 and Topic 2 is 0.895, indicating that dual carbon targets and implementation pathways such as energy conservation, energy saving, and energy intensity constraints frequently co-occur in policy texts.

In contrast, several topic pairs show relatively low similarity, suggesting that their semantics are more specialized and share fewer common elements with mainstream policy narratives. For instance, Topic 27 and Topic 29 exhibit the lowest similarity (0.346), followed by Topic 20–Topic 29 (0.383), Topic 9–Topic 27 (0.404), and Topic 22–Topic 29 (0.430). These low-relatedness pairs typically correspond to more sector-specific or scenario-specific policy expressions, where texts concentrate on institutional links or application contexts and therefore exhibit weaker semantic connections with other topics.

To further assess topic centrality, the average similarity of each topic to all other topics was calculated, and the overall median value of 0.649 was used as a threshold for determining centrality [48]. Topics with an average similarity above this threshold are considered semantically central, including Topic 0 (0.755), Topic 2 (0.724), Topic 1 (0.721), and Topic 11 (0.704). This indicates that these topics are more strongly coupled with most other topics and tend to represent more general governance tasks. By contrast, topics with average similarity below the threshold exhibit stronger semantic independence and can be regarded as relatively peripheral topics, including Topic 29 (0.556), Topic 27 (0.564), Topic 9 (0.585), and Topic 20 (0.591).

4.4.2. Topic Spatial Distribution

To visualize the relative positions and concentration levels of topics in the semantic space, topic vectors were projected into two dimensions, and a topic space bubble map was generated, as shown in Figure 11. Each bubble represents a topic, and bubble size is positively associated with the frequency of that topic across the full set of policy texts. The distance between bubbles is determined by semantic relatedness between topics, where a shorter distance indicates higher semantic similarity.

From the overall distribution, topics exhibit a clear clustered structure in the semantic space, with three core thematic groups. This clustering pattern suggests that China's environmental regulation agenda for air pollution is not evenly dispersed across diverse issues. Instead, it concentrates around several relatively stable policy discourses and instrument systems that form high-density semantic neighborhoods. Beyond these core groups, a small number of topics are located at a noticeable distance from the central areas, forming peripheral topic groups.

The first core thematic group is in the upper high-density region of the figure, which primarily focuses on improving air quality and comprehensive governance of traditional pollution sources. Centered on Topic 0, several topics, including Topic 7, Topic 10, Topic 12, Topic 18, Topic 21, and Topic 25, are tightly distributed around it. This pattern indicates a high level of semantic consistency in the corresponding policy narratives, where multi-source control measures related to coal combustion and dust, campaign-style actions, and regulatory enforcement instruments are frequently articulated together.

The second core thematic group is located in the mid-right area of the figure and reflects policy issues related to structural abatement and green transition. This group aggregates topics such as Topic 1, Topic 2, Topic 4, Topic 5, Topic 6, and Topic 8, covering the dual carbon targets, energy conservation and energy-intensity constraints, cleaner production, accounting and verification arrangements, and renewable energy development.

The formation of this core group suggests that governance objectives have expanded from single pollutant control to encompass synergistic pollution and carbon reduction, as well as broader energy and industrial transition.

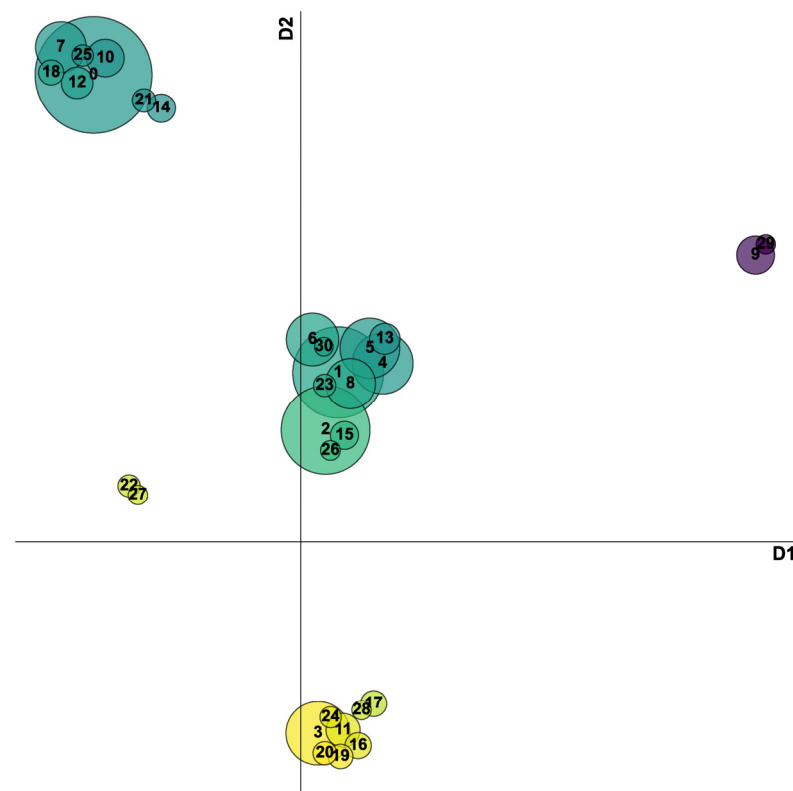


Figure 11. Topic spatial distribution in the semantic space.

The third core thematic group is located in the lower-right high-density region of Figure 10 and mainly corresponds to governance capacity modernization and institutional supporting instruments. This group aggregates topics such as Topic 3, Topic 11, Topic 16, Topic 17, Topic 19, Topic 20, and Topic 28, focusing on institutional elements including monitoring data, permitting management, environmental impact constraints, credit-based regulation, and market-oriented mechanisms. The underlying logic is to enhance measurability, accountability, and risk-sharing capacity in regulation by relying on data infrastructures and market instruments.

In addition to the three core thematic groups, Figure 10 also shows two peripheral topic sub-groups, characterized by fewer bubbles and larger distances from the clusters. This indicates that these topics tend to appear in policy narratives in more specific institutional settings or sectoral scenarios. For example, Topics 9 and 29 are positioned relatively farther from the center of the group, indicating that more sector-specific policy narratives, such as renewable energy subsidies and building energy efficiency, retain a certain degree of semantic independence. Likewise, Topics 22 and 27 form a small cluster on the outer margin of the right group. They are more closely related to scenario-based issues and pollution-related damages and thus remain relatively independent in semantic terms.

To clarify the internal relationships among the thematic groups, Table 5 presents the inter- and intra-group cosine similarities. The diagonal values represent intra-group similarities, indicating internal cohesion, while the off-diagonal values capture inter-group interactions. All groups exhibit strong internal semantic density with intra-group similarities exceeding 0.69, and the two peripheral groups show the highest internal cohesion at over 0.81.

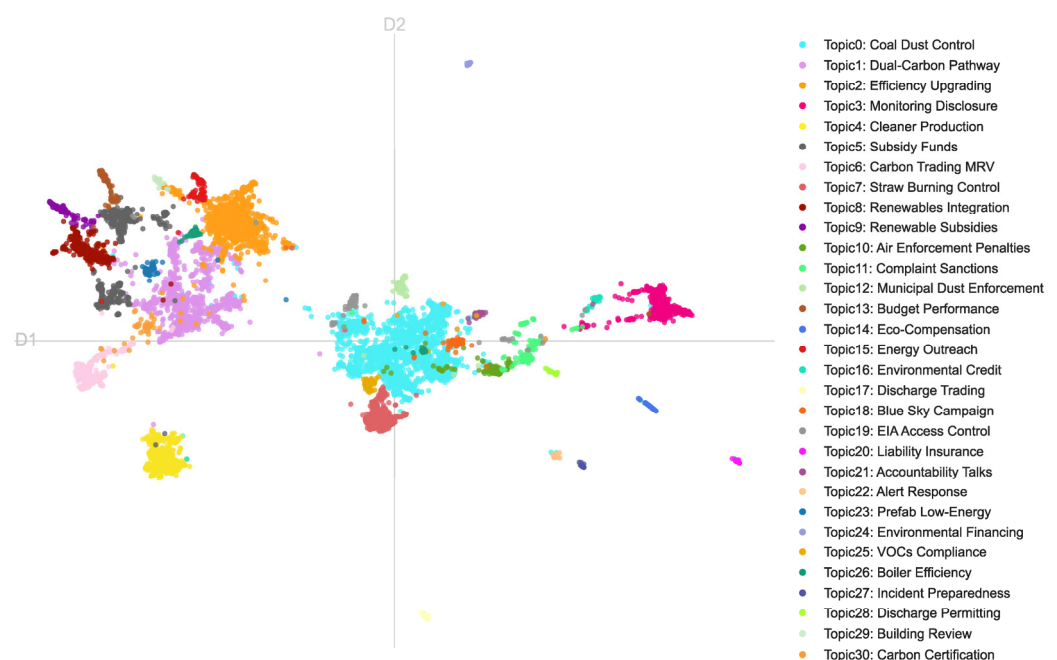
Table 5. The inter-group similarity and intra-group similarity of the five theme groups.

Topic Group	Core Group1	Core Group 2	Core Group 3	Peripheral Group 1	Peripheral Group 2
Core group 1	0.794	0.662	0.672	0.546	0.649
Core group 2	0.662	0.722	0.612	0.655	0.539
Core group 3	0.672	0.612	0.699	0.490	0.644
Peripheral group 4	0.546	0.655	0.490	0.816	0.407
Peripheral group 5	0.649	0.539	0.644	0.407	0.818

The fundamental distinction between core and peripheral themes lies in their inter-group interactions. The three core groups maintain robust cross-correlations with one another, evidenced by inter-group similarities ranging from 0.612 to 0.672. In contrast, the peripheral groups display significantly lower similarities with other groups. For instance, the inter-group similarity between the two peripheral groups drops to 0.407. This structural contrast justifies their classification: core groups represent general strategic agendas, whereas peripheral groups act as highly specialized, relatively isolated policy instruments.

4.4.3. Document Spatial Distribution

To examine the spatial separability and internal coherence of the topic partition at the document level, the document embedding vectors were projected into two dimensions, and the distribution of policy texts in the topical space was visualized, as shown in Figure 12. Each point represents a single policy document, and the distance between points reflects semantic similarity. Point colors indicate topic assignments.

**Figure 12.** Document topic distribution atlas.

In terms of clustering intensity, topics display clear differences in compactness in the two-dimensional space. Documents assigned to Topic 9, Topic 10, Topic 7, Topic 4, Topic 8, and Topic 6 form comparatively compact, same-colored clusters or localized high-density point clouds, suggesting stronger semantic consistency within these topics and more stable terminological systems and governance contexts in the corresponding policy texts. By contrast, Topics 0, 1, and 2 cover a wider area and exhibit a more dispersed distribution, indicating that these topics are more likely to carry comprehensive narratives or

cross-domain policy expressions and therefore contain more mixed information involving multiple governance targets and instrument combinations. Topic 5 shows even stronger dispersion, suggesting that texts concerning fiscal funding, rewards and subsidies, and project-based implementation are often embedded in diverse governance scenarios, which makes the semantic boundaries of this topic more prone to spillover.

In terms of boundary clarity, the topics exhibit relatively strong spatial separability, although the degree of separation varies across topics. For example, Topics 3, 4, 6, and 7 display comparatively clear document-level boundaries, with a lower proportion of heterogeneously colored points mixed among the same-colored clusters. This indicates that their semantic boundaries against other topics are relatively distinct, allowing documents to be assigned to these topics with greater stability. Meanwhile, several topics exhibit relatively blurred boundaries, as evidenced by the substantial mixing of differently colored points along cluster margins or the continuous spillover of the same-colored points into neighboring regions. Topic 0 is one of the high-frequency topics with more pronounced boundary spillover, reflecting its capacity to absorb and connect multiple pollution source governance elements and diverse regulatory instruments within comprehensive policy narratives.

Regarding inter-topic overlap, Figure 12 shows several stable overlap zones. Topics 0 and 7 are spatially close and form a clear mixed region, suggesting that comprehensive control narratives frequently co-occur with scenario-based governance measures such as straw burning and dust control within numerous policy texts. Overlaps are even more evident among institution-oriented topics: Topics 10 and 11 exhibit strong interpenetration, and Topic 11 also substantially overlaps with Topic 3. This implies that enforcement and penalties, public reporting, and monitoring and data measures are often articulated jointly in policy texts, forming an interconnected sequence from information discovery to evidentiary support and enforcement correction. The presence of overlap zones does not undermine the validity of topic identification. It reveals that multiple policy instruments and governance stages are frequently combined within the same document, thereby generating natural transitional areas in the semantic space.

4.5. Topic Dynamic Analysis

Dynamic topic analysis is used to characterize how policy issues change over time. This study calculates the annual frequency of policy texts associated with each topic and generates topic-evolution curves, as shown in Figure 13. Because topics differ substantially in the magnitude of fluctuations, the analysis focuses on trajectories that show pronounced changes or are particularly representative, in order to reveal how the policy agenda evolves across governance stages.

Overall, topic trajectories exhibit clear stage-related features. First, the aggregate pattern can be summarized as a trajectory of rapid early increase, a high level during the intensive governance phase, followed by a decline and a subsequent rebound. Topic 0 provides a representative example. It increased markedly to 263 documents in 2013 and peaked at 354 documents in 2018, then declined to 61 documents in 2022, but rebounded to 163 documents in 2023. This indicates that the comprehensive governance agenda, oriented toward air quality improvement, has remained central over the long term, while its textual intensity fluctuates by stage in response to adjustments in the governance rhythm. By contrast, scenario-based governance topics display a more “short-term concentrated surge” pattern. For instance, Topic 7 rose sharply to 120 documents in 2018 and remained at 111 documents in 2019, then quickly declined to 16 documents in 2021 and 10 documents in 2022. This suggests that issues such as straw-burning bans and dust control tend to generate intensive policy releases during special campaign periods, after which governance shifts toward routine management.

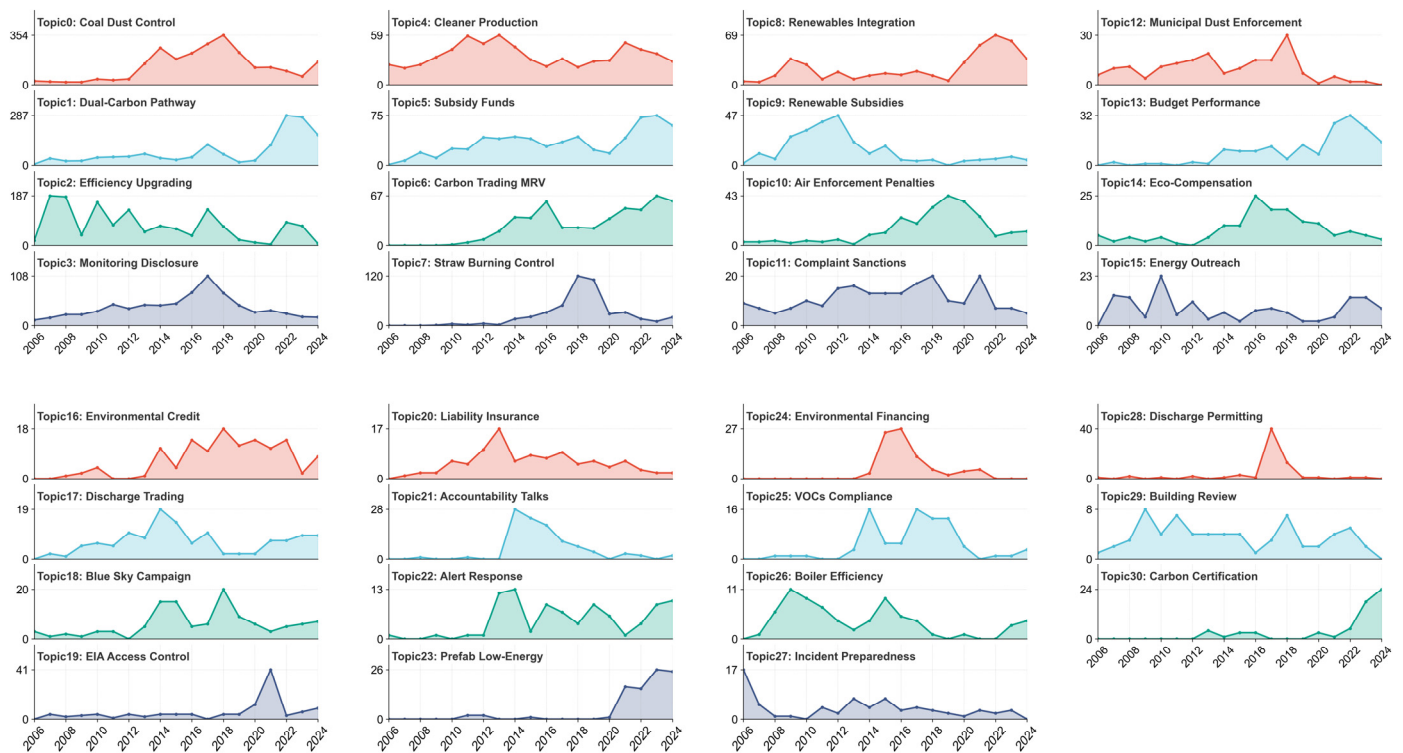


Figure 13. Temporal evolution in annual topic frequencies from 2006 to 2024.

In comparison, low-carbon-transition-related issues show a structural rise in the later years of the sample. Topic 1 increased from 17 documents in 2019 to 143 documents in 2020, peaked at 287 documents in 2021, and remained high at 277 documents in 2022. This indicates that the dual carbon goals have been incorporated into air pollution governance policies in recent years, marking a period of high-intensity policy attention. In addition, Topic 8 rose from a low level to 86 documents in 2020 and remained relatively high in 2021–2022, indicating that, beyond conventional pollution-source control, policy texts increasingly emphasize upstream abatement through energy structure optimization and system-level substitution. Topic 6 increased to 86 documents in 2020 and remained at around 60 documents in 2022–2023, suggesting that, with the improvement of relevant institutional systems and stronger compliance requirements, expressions related to accounting, verification, greenhouse gases, and emission allowance management have increased markedly.

Instrument-oriented topics exhibit more pronounced temporal cyclicity. Topic 10 peaked at 63 documents in 2020 but dropped sharply to 8 documents in 2021 and remained low thereafter, indicating that command-and-control instruments may be intensified in specific periods and then shift toward routine enforcement. By contrast, the fiscal funding and incentive Topic 5 remained at a high level during 2020–2022, with 59, 72, and 75 documents, respectively, and remained at 60 documents in 2023, suggesting that incentive-based instruments have been applied more frequently in the later stages of environmental governance.

5. Conclusions

5.1. Key Findings

Based on air pollution environmental regulation policy texts issued by the central government and 31 provincial-level governments, this study constructs a policy corpus and applies BERTopic for semantic topic identification. By integrating keyword weights, hierarchical structures, similarity, spatial distribution, and dynamic topic analysis, the

study characterizes the main content and evolutionary patterns of the policy agenda. The main conclusions can be summarized in four points.

First, topic content identification indicates a clear core topic structure in the policy agenda. Topic 0–Topic 11, each with more than 200 documents, jointly cover 9625 texts, accounting for 82.8% of the full sample. Among them, Topic 0 is the largest topic, accounting for 21.3%, reflecting concentrated attention to coal control, dust governance, and regulatory enforcement under an air quality improvement orientation. Topic 1 and Topic 2 correspond to target-oriented narratives and pathway-oriented instruments, respectively, and they constitute important components of the policy agenda. Meanwhile, topics such as Topic 3, Topic 5, Topic 6, and Topic 8 indicate the sustained presence of institutional support and implementation instruments in policy texts.

Second, hierarchical clustering reveals stable combinations among topics. Topic 0 merges with Topic 18 and Topic 7 at relatively small distances, suggesting that comprehensive governance narratives often co-occur with campaign style action expressions and non-point source governance requirements. Topics 1 and 2 merge at an early stage, indicating that low carbon transition goals are frequently articulated together with pathway instruments, such as energy conservation and energy intensity constraints, in policy texts. The close adjacency between Topic 10 and Topic 11 reflects the joint use of enforcement penalties and reporting-based oversight in institutional arrangements.

Third, topic distance and similarity results indicate a structural pattern of “core-topic concentration and specialized-topic independence”. The topic bubble map and document projection indicate that topics form three relatively concentrated groups in the semantic space, corresponding to comprehensive control issues related to air quality, pollution, carbon reduction synergy, and transition-related issues, as well as monitoring and rule-based institutional issues. The similarity heatmap further reveals that some topic pairs exhibit high similarity, such as Topic 0–Topic 7 (0.944), while other pairs show low similarity, as seen in Topic 27–Topic 29 (0.346). In the document projection, compact clusters and overlap regions coexist, suggesting that certain topics have strong co-occurrence relationships in policy texts.

Fourth, dynamic topic analysis demonstrates clear stage-wise changes in the policy agenda. The comprehensive control Topic 0 peaked at 354 documents in 2018, declined to 61 documents in 2022, and rebounded to 163 documents in 2023, indicating that the intensity of comprehensive governance-related policy issuance fluctuates with governance stages. The scenario-based Topic 7 recorded 120 documents in 2018 and 111 documents in 2019, but declined to 10 documents in 2022, reflecting the stage-specific nature of campaign-oriented governance issues. The low-carbon transition Topic 1 increased markedly in the later period, from 17 documents in 2019 to 287 documents in 2021, and it remained at 277 documents in 2022, indicating a significant strengthening of the pollution carbon reduction synergy in recent years.

5.2. Limitations

Despite its contributions, this study has several limitations. First, the spatial scope is primarily restricted to central and provincial-level policies, which may not fully capture the micro-scale variations at the municipal or county levels. Second, while the BERTopic model effectively identifies explicit thematic patterns, it has a limited capacity to analyze the more implicit strategic patterns often embedded in political discourse. Most importantly, it focuses on the semantic content of policy texts and does not yet link topics to quantitative indicators such as air quality improvements or emission reduction performance for causal identification or effectiveness evaluation.

Future research should address these gaps by expanding the analytical framework in two primary directions. Spatially, incorporating policy texts from city and sector levels would enable a more comprehensive multi-scalar analysis of China's environmental governance hierarchy. Empirically, the priority is to transition from thematic analysis to impact evaluation. By utilizing the topic intensities identified in this study as independent variables, future work can employ econometric methods, such as difference-in-differences or panel data models, to identify the causal impact of regulatory themes on ecological performance. Additionally, incorporating international policy would facilitate comparative studies to examine commonalities and differences across diverse governance systems.

5.3. Policy Implications

Based on the above findings, three policy implications are proposed. First, the comprehensive governance policy system, with a particular emphasis on air quality, should be continually strengthened. Given that the comprehensive control topic remains dominant and shows a rebound in the later period, stable governance requirements should be maintained for key links such as coal-related emissions and dust in priority regions and during key periods. Furthermore, policymakers should prioritize the transition from campaign-style enforcement to a long-term institutional framework by embedding air quality benchmarks into routine administrative audits, thereby ensuring the stability of regulatory intensity across different economic cycles.

Second, the synergy between pollution reduction and carbon reduction should be translated into actionable policy packages and aligned with monitoring, reporting, and verification arrangements [49]. Considering the strong linkage between low-carbon transition themes and energy conservation and energy intensity constraints, coordinated policy design should be promoted in key sectors and regions by combining instruments such as energy intensity constraints, energy efficiency improvements, cleaner production, and renewable energy substitution [50]. Meanwhile, accounting-, verification-, and compliance-related rules should be improved to enhance the measurability and assessability of emission reduction targets.

Third, the implementation efficiency of institutional support instruments and scenario-based governance should be enhanced. For institution-oriented tools such as environmental monitoring and reporting-based supervision, mechanisms for data quality control, information sharing, and feedback should be improved to support enforcement and evaluation [51]. For scenario-based issues such as straw burning and dust control, standardized routine management procedures and accountability mechanisms should be established to reduce the risks associated with a rapid decline in governance intensity after the end of special campaigns.

Author Contributions: Conceptualization, Y.X. and Y.W.; methodology, Y.X.; software, Y.X.; validation, Y.X., Y.W.; formal analysis, Y.X.; investigation, Y.X.; resources, Y.W.; data curation, Y.W.; writing—original draft preparation, Y.X.; writing—review and editing, Y.X.; visualization, Y.X.; supervision, Y.W.; project administration, Y.W.; funding acquisition, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Jiangxi Provincial Social Science “14th Five-Year Plan” Fund Project (22GL39) and the Key Research Base Project of Jiangxi Provincial Philosophy and Social Sciences (24ZXSKJD01).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset available on request from the corresponding author. The raw data supporting the conclusions of this article will be made available by the corresponding author on request.

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

References

1. Dong, F.; Wang, Y.; Su, B.; Hua, Y.; Zhang, Y. The process of peak CO₂ emissions in developed economies: A perspective of industrialization and urbanization. *Resour. Conserv. Recycl.* **2019**, *141*, 61–75. [\[CrossRef\]](#)
2. Yin, P.; Brauer, M.; Cohen, A.J.; Wang, H.; Li, J.; Burnett, R.T.; Stanaway, J.D.; Causey, K.; Larson, S.; Godwin, W.; et al. The effect of air pollution on deaths, disease burden, and life expectancy across China and its provinces, 1990–2017: An analysis for the Global Burden of Disease Study 2017. *Lancet Planet Health* **2020**, *4*, e386–e398. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Lu, X.; Zhang, S.; Xing, J.; Wang, Y.; Chen, W.; Ding, D.; Wu, Y.; Wang, S.; Duan, L.; Hao, J. Progress of air pollution control in China and its challenges and opportunities in the ecological civilization era. *Engineering* **2020**, *6*, 1423–1431. [\[CrossRef\]](#)
4. Xu, M.; Qin, Z.; Zhang, S. Integrated assessment of cleaning air policy in China: A case study for Beijing-Tianjin-Hebei region. *J. Clean. Prod.* **2021**, *296*, 126596. [\[CrossRef\]](#)
5. Bu, C.; Shi, D. The emission reduction effect of daily penalty policy on firms. *J. Environ. Manag.* **2021**, *294*, 112922. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Feng, Y.; Cheng, C.; Hu, S.; Cao, A. Campaign-style governance of air pollution in China? A comprehensive analysis of the central environmental protection inspection. *Front. Public Health* **2023**, *11*, 1081573. [\[CrossRef\]](#)
7. Zhang, Q.; Yin, Z.; Lu, X.; Gong, J.; Lei, Y.; Cai, B.; Cai, C.; Chai, Q.; Chen, H.; Dai, H. Synergetic roadmap of carbon neutrality and clean air for China. *Environ. Sci. Ecotechnol.* **2023**, *16*, 100280. [\[CrossRef\]](#)
8. Wen, Z.; Ma, X.; Xu, W.; Si, R.; Liu, L.; Ma, M.; Zhao, Y.; Tang, A.; Zhang, Y.; Wang, K.; et al. Combined short-term and long-term emission controls improve air quality sustainably in China. *Nat. Commun.* **2024**, *15*, 5169. [\[CrossRef\]](#)
9. Li, Z.; Zhao, K.; Yuan, X.; Zhou, Y.; Yang, L.; Geng, H. Evolution and Control of Air Pollution in China over the Past 75 Years: An Analytical Framework Based on the Multi-Dimensional Urbanization. *Atmosphere* **2024**, *15*, 1093. [\[CrossRef\]](#)
10. Feng, T.; Sun, Y.; Shi, Y.; Ma, J.; Feng, C.; Chen, Z. Air pollution control policies and impacts: A review. *Renew. Sustain. Energy Rev.* **2024**, *191*, 114071. [\[CrossRef\]](#)
11. Jiang, X.; Li, G.; Fu, W. Government environmental governance, structural adjustment and air quality: A quasi-natural experiment based on the Three-year Action Plan to Win the Blue Sky Defense War. *J. Environ. Manag.* **2021**, *277*, 111470. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Cheng, J.; Yi, J.; Dai, S.; Xiong, Y. Can low-carbon city construction facilitate green growth? Evidence from China's pilot low-carbon city initiative. *J. Clean. Prod.* **2019**, *231*, 1158–1170. [\[CrossRef\]](#)
13. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794.
14. Du, K.; Cheng, Y.; Yao, X. Environmental regulation, green technology innovation, and industrial structure upgrading: The road to the green transformation of Chinese cities. *Energy Econ.* **2021**, *98*, 105247. [\[CrossRef\]](#)
15. Li, J.; Huang, J.; Li, B. Do command-and-control environmental regulations realize the win-win of “pollution reduction” and “efficiency improvement” for enterprises? Evidence from China. *Sustain. Dev.* **2024**, *32*, 3271–3292. [\[CrossRef\]](#)
16. Ouyang, X.; Li, Q.; Du, K. How does environmental regulation promote technological innovations in the industrial sector? Evidence from Chinese provincial panel data. *Energy Policy* **2020**, *139*, 111310. [\[CrossRef\]](#)
17. Ge, T.; Hao, X.; Li, J. Effects of public participation on environmental governance in China: A spatial Durbin econometric analysis. *J. Clean. Prod.* **2021**, *321*, 129042. [\[CrossRef\]](#)
18. Yan, Z.; Yu, Y.; Du, K.; Zhang, N. How does environmental regulation promote green technology innovation? Evidence from China's total emission control policy. *Ecol. Econ.* **2024**, *219*, 108137. [\[CrossRef\]](#)
19. Lin, J.; Long, C.; Yi, C. Has central environmental protection inspection improved air quality? Evidence from 291 Chinese cities. *Environ. Impact Assess. Rev.* **2021**, *90*, 106621. [\[CrossRef\]](#)
20. Jiang, J.; Zhang, Q.; Hui, Y. The impact of market and non-market-based environmental policy instruments on firms' sustainable technological innovation: Evidence from Chinese firms. *Sustainability* **2023**, *15*, 4425. [\[CrossRef\]](#)
21. Tian, B.; Yu, J.; Tian, Z. The impact of market-based environmental regulation on corporate ESG performance: A quasi-natural experiment based on China's carbon emission trading scheme. *Heliyon* **2024**, *10*, e26687. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Lan, L.; Duan, Y.; Zhou, Z.; Zhong, S. Voluntary Environmental Regulation and Corporate Green Innovation: Evidence From the “Green Factory” Certification. *Bus. Strateg. Environ.* **2025**. [\[CrossRef\]](#)
23. Wang, N.; Yang, G.; Wu, J.; Wenya, L. Impact of environmental information disclosure on green technology innovation: Evidence from China's Listed Companies. *Financ. Res. Lett.* **2025**, *71*, 106375. [\[CrossRef\]](#)
24. Wu, W.; Huang, M.; Lv, H.; Wang, W.; Zhang, M. Does the air pollution joint prevention and control policy possess the Porter effect in China-evidence from two perspectives. *Environ. Impact Assess. Rev.* **2021**, *91*, 106678. [\[CrossRef\]](#)

25. Liu, Y.; Liu, M.; Wang, G.; Zhao, L.; An, P. Effect of Environmental Regulation on High-quality Economic Development in China—An Empirical Analysis Based on Dynamic Spatial Durbin Model. *Environ. Sci. Pollut. Res.* **2021**, *28*, 54661–54678. [\[CrossRef\]](#)
26. Wang, C.; Ling, J.; Liu, Y.; Liu, B.; Deng, N. Can air quality ecological compensation improve environmental welfare performance? Based on the “Win–Win–Win” perspective of economy–ecology–welfare. *J. Clean. Prod.* **2025**, *489*, 144604. [\[CrossRef\]](#)
27. Ge, X.; Li, Y.; Yang, H. The green paradox of time dimension: From pilot to national carbon emission trading system in China. *Environ. Impact Assess. Rev.* **2024**, *109*, 107642. [\[CrossRef\]](#)
28. Yang, G.; Zha, D.; Wang, X.; Chen, Q. Exploring the nonlinear association between environmental regulation and carbon intensity in China: The mediating effect of green technology. *Ecol. Indic.* **2020**, *114*, 106309. [\[CrossRef\]](#)
29. Fan, L.; Yao, S.; Zhou, C.; Richardson-Barlow, C. Environmental regulatory policies and local air quality: Evidence from a comparison of environmental policy effects in China. *J. Clean. Prod.* **2025**, *518*, 145906. [\[CrossRef\]](#)
30. Liu, M.; Buntaine, M.T.; Anderson, S.E.; Zhang, B. Transparency by Chinese cities reduces pollution violations and improves air quality. *Proc. Natl. Acad. Sci. USA* **2025**, *122*, e2406761122. [\[CrossRef\]](#)
31. Wang, X.; Zhao, H.; Zhu, H.; Wang, F. Towards intelligent policy analysis: A discourse structure parsing technique for Chinese government document. *Inf. Process. Manag.* **2023**, *60*, 103363. [\[CrossRef\]](#)
32. Yang, Y.; Tan, X.; Shi, Y.; Deng, J. What are the core concerns of policy analysis? A multidisciplinary investigation based on in-depth bibliometric analysis. *Humanit. Soc. Sci. Commun.* **2023**, *10*, 190. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Gyódi, K.; Nawaro, Ł.; Paliński, M.; Wilamowski, M. Informing policy with text mining: Technological change and social challenges. *Qual. Quant.* **2023**, *57*, 933–954. [\[CrossRef\]](#)
34. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
35. Nellutla, A.P.; Hudnurkar, M.; Ambekar, S.S.; Lidbe, A.D. Online Product Reviews and Their Impact on Third Party Sellers Using Natural Language Processing. *Int. J. Bus. Intell. Res.* **2021**, *12*, 26–47. [\[CrossRef\]](#)
36. Li, T.; Ke, X.; Shi, H. Topic modeling and evolutionary trends of China’s language policy: A LDA-ARIMA approach. *PLoS ONE* **2025**, *20*, e0324644. [\[CrossRef\]](#)
37. Hankar, M.; Kasri, M.; Beni-Hssane, A. A comprehensive overview of topic modeling: Techniques, applications and challenges. *Neurocomputing* **2025**, *628*, 129638. [\[CrossRef\]](#)
38. Feng, Y.; Chen, Z.; Zhang, Y.; Huang, W.; Zhang, X.; He, S. BERTopic_Teen: A multi-module optimization approach for short text topic modeling in adolescent health. *Front. Public Health* **2025**, *13*, 1608241. [\[CrossRef\]](#)
39. Ma, L.; Chen, R.; Ge, W.; Rogers, P.; Lyn-Cook, B.; Hong, H.; Tong, W.; Wu, N.; Zou, W. AI-powered topic modeling: Comparing LDA and BERTopic in analyzing opioid-related cardiovascular risks in women. *Exp. Biol. Med.* **2025**, *250*, 10389. [\[CrossRef\]](#)
40. Han, X. Evolution of research topics in LIS between 1996 and 2019: An analysis based on latent Dirichlet allocation topic model. *Scientometrics* **2020**, *125*, 2561–2595. [\[CrossRef\]](#)
41. Farea, A.; Tripathi, S.; Glazko, G.; Emmert-Streib, F. Investigating the optimal number of topics by advanced text-mining techniques: Sustainable energy research. *Eng. Appl. Artif. Intell.* **2024**, *136*, 108877. [\[CrossRef\]](#)
42. Roberts, M.E.; Stewart, B.M.; Tingley, D. stm: An R Package for Structural Topic Models. *J. Stat. Softw.* **2019**, *91*, 1–40. [\[CrossRef\]](#)
43. Dybowski, T.P.; Adämmer, P. The economic effects of US presidential tax communication: Evidence from a correlated topic model. *Eur. J. Political Econ.* **2018**, *55*, 511–525. [\[CrossRef\]](#)
44. Zhu, Y.; Li, G. Analyzing the evolution of environmental policy themes and governance stakeholder networks: A computational approach. *Int. J. Inf. Manag. Data Insights* **2025**, *5*, 100382. [\[CrossRef\]](#)
45. Xue, W.; Wang, L.; Yang, Z.; Xiong, Z.; Li, X.; Xu, Q.; Cai, Z. Can clean heating effectively alleviate air pollution: An empirical study based on the plan for cleaner winter heating in northern China. *Appl. Energy* **2023**, *351*, 121923. [\[CrossRef\]](#)
46. Zhou, X.; Cao, G.; Peng, B.; Xu, X.; Yu, F.; Xu, Z.; Yan, Y.; Du, H. Citizen environmental complaint reporting and air quality improvement: A panel regression analysis in China. *J. Clean. Prod.* **2024**, *434*, 140319. [\[CrossRef\]](#)
47. Huang, C.; Fu, S.C.; Chan, K.C.; Wu, C.; Chao, C.Y.H. Evaluating China’s 2030 carbon peak goal: Post-COVID-19 systematic review. *Renew. Sustain. Energy Rev.* **2025**, *209*, 115128. [\[CrossRef\]](#)
48. Rombach, M.P.; Porter, M.A.; Fowler, J.H.; Mucha, P.J. Core-Periphery Structure in Networks. *Siam J. Appl. Math.* **2012**, *74*, 167–190. [\[CrossRef\]](#)
49. Fan, X.; Zhou, Y.; Xie, Q. Assessment on the synergistic effect of pollution and carbon reductions in low-carbon city pilot policy: Based on effectiveness and efficiency perspectives. *Environ. Dev. Sustain.* **2025**, *27*, 12945–12969. [\[CrossRef\]](#)

50. Tao, M.; Wen, L.; Sheng, M.S.; Yan, Z.J.; Poletti, S. Dynamics between energy intensity and carbon emissions: What does the clustering effect of labor and capital play? *J. Clean. Prod.* **2024**, *452*, 142223. [[CrossRef](#)]
51. Zhao, J.; Tang, Y.; Zhu, X.; Zhu, J. National environmental monitoring and local enforcement strategies. *Nat. Cities* **2025**, *2*, 58–69. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.