

# scientific data



**OPEN**  
DATA DESCRIPTOR

## Climate risk expressions of public views and government responses in China

Xiaolei Sun<sup>1,2</sup>, Longfei Li<sup>1,2</sup>, Yiran Shen<sup>1,2</sup>, Yingjie Sheng<sup>1,2</sup>, Dayong Zhang<sup>3</sup>✉ & Qiang Ji<sup>1,2</sup>✉

Expressions of public views on climate risk and corresponding government responses serve as crucial indicators of the dynamic interaction between the public and the government in tackling climate change. Focusing on the context of China, this study utilizes word embedding techniques to construct indices of expressions of public views and government responses on climate risk at the national, provincial, and city levels for the first time. These indices are derived from data collected on the *Message Board for Leaders* platform spanning from January 2011 to December 2023. The developed indices not only facilitate the study of expressions of public views and government responses to climate risk but also provide robust data support for enhancing climate risk communication between the government and the public. Furthermore, these indices are instrumental in bolstering regional climate adaptation initiatives and accelerating the development of climate-resilient cities.

### Background and Summary

Extreme climate events are characterized by their high destructiveness and unpredictability<sup>1</sup>. The frequent occurrence of these events has a significant impact on residents' lives, leading to widespread expressions of public views<sup>2</sup>. These expressions are conveyed through various feedback channels, forming a dynamic system of expression feedback. If these expressions are not effectively responded, their long-term accumulation may potentially pose a threat to local social stability and order. Therefore, establishing an efficient mechanism for identifying and responding to expressions related to climate risk is crucial for accurately assessing and mitigating climate risks, as well as for ensuring social harmony and stability<sup>3</sup>.

In the context of expressions of public views generated by climate events, government measures aimed at protecting and adapting to these expressions can improve societal climate adaptation effectiveness and foster public confidence in the community's ability to cope with climate threats<sup>4</sup>. The public derives an understanding of regional climate risk trends from the government's handling of specific risk events in particular locations, and through a long-term process of assessment and cognition, they develop an understanding of the region's risk adaptation capacity<sup>5</sup>. Real-time government responses to expressions of public views about climate risk—government attitudes, reactions, and communication—help reduce information discrepancies in climate risk communication, thereby strengthening regional adaptation capabilities<sup>6</sup>. This interactive mechanism between government climate responses and expressions of public views is crucial to providing people-centered climate adaptation actions<sup>7</sup>. The key elements in measuring the intensity of expressions of public views under climate risk are the public's concerns about urgency and the degree of loss. The obstacles to measuring these two aspects are as follows.

The first challenge is related to sentiment—the urgency of climate events emphasizes the public's sentimental sensitivity to risk. In existing research, both survey questionnaires and social media data have served as valuable data sources for collecting sentimental information<sup>8</sup>. Previous studies have measured trends in emotional changes by analyzing social media posts on platforms such as *Twitter* and *Weibo*<sup>9</sup>, providing a research paradigm based on social text data. However, social media platforms often contain noise<sup>10</sup> and echo chamber effects<sup>11</sup>, which may distort or exaggerate sentimental. Thus, accurately assessing event urgency and capturing sentiment require social text data with lower bias.

<sup>1</sup>Institutes of Science and Development, Chinese Academy of Sciences, Beijing, China. <sup>2</sup>School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing, China. <sup>3</sup>Research Institute of Economics and Management, Southwestern University of Finance and Economics, Chengdu, China. ✉e-mail: [dzhang@swufe.edu.cn](mailto:dzhang@swufe.edu.cn); [jqwxnjq@163.com](mailto:jqwxnjq@163.com)

The second challenge pertains to loss, especially the public's assessment of the severity of losses caused by climate events. In existing studies, questionnaires have been widely used to collect data related to the public's assessment of losses from specific climate events<sup>8,12</sup>. Proxy variables such as coastal housing sales data<sup>13</sup> and weather forecast data<sup>14</sup> have also been used to infer the public's perception of losses. This study aims to provide timely data support for climate risk communication by quantifying expressions of public views regarding climate events and corresponding governmental responses. To this end, real-time social media data are used to dynamically capture the public's discussions on losses<sup>10,15</sup> and quantify the public's perception of climate risk losses. The difficulty lies in how to extract key climate risk terms related to the loss statements in public views from social media data. On the one hand, the lists of climate risk terms summarized in existing studies, such as global warming<sup>16</sup>, corporate social responsibility<sup>17</sup>, and climate policy uncertainty<sup>18</sup>, may overlook many specific words used by the public to describe climate risks and are difficult to comprehensively summarize loss concerns. On the other hand, as a typical large text corpus, social media data poses challenges in identifying all relevant key terms, which requires the construction of a specialized dictionary tailored to the research question<sup>19</sup>. Therefore, it is necessary to integrate the public's expressions, jointly consider climate characteristic information and risk concern information, and construct a dictionary of climate risk losses.

To capture more authentic sentiment features, text data from the *Message Board for Leaders* (<https://liuyan.people.com.cn/>) are utilized to extract public sentiment information. It is an authoritative platform dedicated to collecting expressions of public views that is hosted by the official media *People's Daily Online* in China. The registration process is unrestricted, and users can post messages through real-name authentication to ensure the authenticity and credibility of the information, while also allowing anonymous posting to protect user privacy. It primarily reflects the genuine expressions of the public, capturing the difficulties and challenges citizens face during daily life and disaster scenarios with only minimal presence of advertisements and false information<sup>20</sup>. Since individual posts their own expressions without the process of reposting others' expressions, the echo chamber effect is significantly lower than on social media platforms like *Weibo* and *Facebook*<sup>21</sup>. Furthermore, since local officials do not have the authority to censor or selectively filter posts, the platform encourages truthful reporting from citizens, with less external scrutiny<sup>22</sup>.

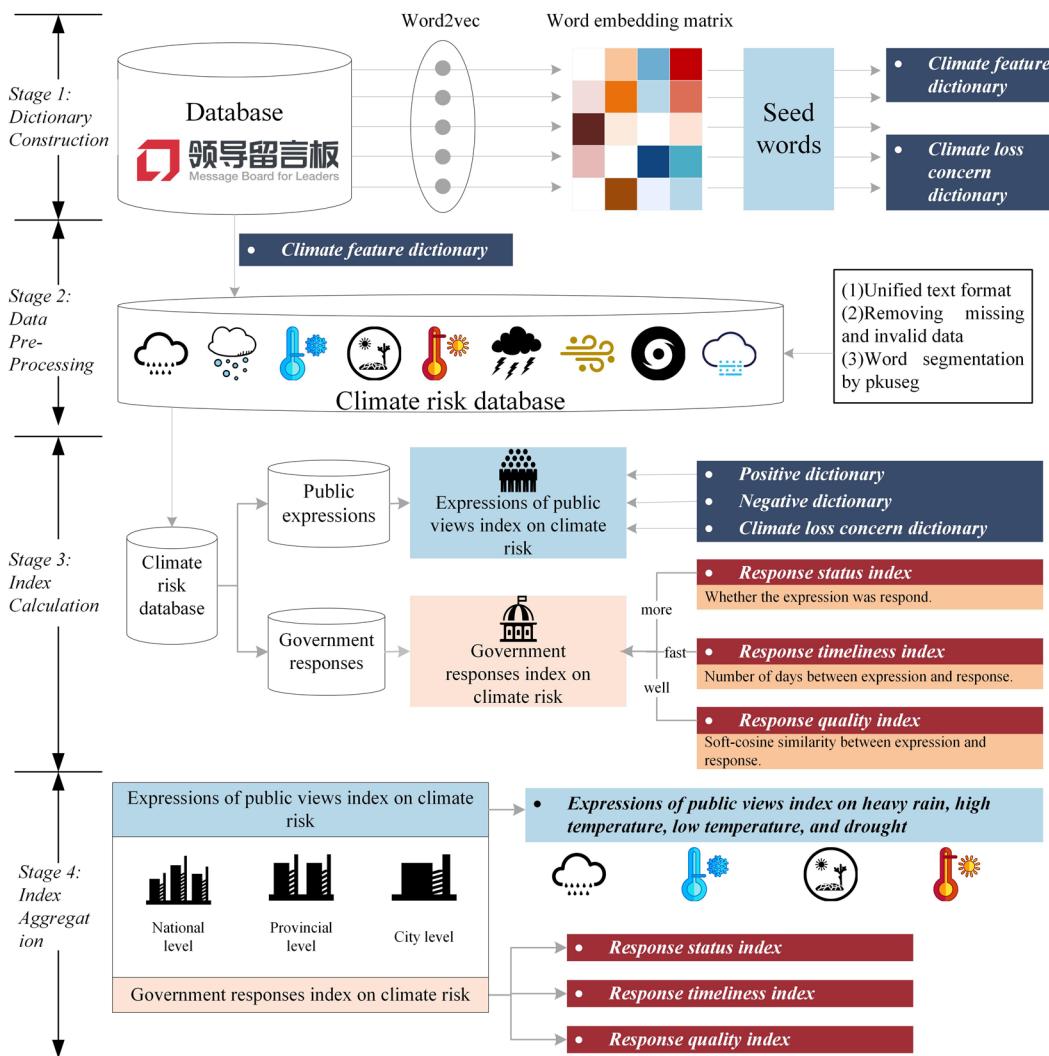
AI for social sciences and humanities has emerged as a new research paradigm in the fields of social sciences and humanities, reshaping the research framework of public administration and decision-making. To extract the public's concerns about the severity of losses caused by climate events, this study utilizes word embedding methods to identify collectively representative expressions of severity. Based on co-occurrence data from the *Message Board for Leaders*, this study calculates words that are simultaneously close to both climate feature information and risk loss concern information in the vector space, thereby constructing a climate loss concern dictionary. This approach enables the joint processing of climate feature information and loss concern information, allowing the constructed dictionary to accurately capture the public's concerns about the severity of losses related to the specific subject of climate. Additionally, through manually reading and labeling validation, the study confirms that the dictionary effectively captures collective and representative expressions of loss related to climate risks.

For the aforementioned reasons, this study used the *Message Board for Leaders* as the data source. A total of 3,084,890 texts containing risk expressions of public views and government responses were collected between January 2011 and December 2023. From a corpus of 222,923 paired messages and responses, a climate feature dictionary containing 292 words including nine types of climate events and a climate loss concern dictionary containing 1,820 words were extracted to construct the index of public views on climate risk<sup>23</sup>. The proposed methods for identifying climate features and extracting climate loss concerns were validated through manual auditing and spatial distribution analysis, ensuring that the methods can accurately collect the information concerns in the messages. To construct the government response index on climate risk, government responses from the same period and region were extracted and integrated. These responses were evaluated across three dimensions: status, timeliness, and quality. Specifically, the response status index measures the proportion of government responses, the response timeliness index evaluates the speed of responses, and the response quality index assesses the effectiveness and relevance of the responses.

The main contributions of this study are twofold. First, in the context of climate adaptation actions, this study integrates bottom-up expressions of public views characteristics and top-down government response capabilities into the same analytical framework. Inspired by media-based climate change concern indices<sup>24</sup>, this study constructs new expressions of public views and government response indices for climate risk based on the characteristics of the message board text data. The raw climate-related message board data collected in this study may represent one of the first instances where such data is made accessible to the English-speaking research community. Second, the national-level expressions of public views and government response indices for climate risk are extended to the provincial and city levels. This dataset can directly provide climate risk communication information to government decision-makers, offering more precise risk information support for targeted emergency responses to climate events and enhancing communication efficiency with the public. Additionally, the dataset can support researchers in the field of climate governance, contributing to the improvement of regional climate adaptation effectiveness and the development of climate-resilient cities.

## Methods

To construct the indices of expressions of public views and government responses on climate risk, this study extracted relevant climate risk information from the *Message Board for Leaders*. The process comprised four main stages: dictionary construction, data preprocessing, index calculation, and index aggregation. The framework is shown in Fig. 1.



**Fig. 1** Research framework.

**Dictionary construction.** The dictionaries were constructed using the following approach. First, the model training used all the messages and their corresponding responses from January 2011 to December 2023 in the *Message Board for Leaders*. Word2Vec model was trained to better adapt to the linguistic features of public risk expressions, in response to the short-text, high-frequency word interaction characteristics of this corpus. The study estimated word embeddings using the Word2Vec framework. The hidden layer consisted of 300 “hidden neurons” and each word was represented as a 300-dimensional vector, with a context window size of 6. This method maps each word in the corpus to a finite vector space; the semantic similarity between any two words was measured by calculating the cosine distance between their vectors in preparation for subsequent dictionary expansion.

Three main dictionaries were used in this study. Firstly, the climate feature dictionary was composed of keywords that directly describe climate events. Climate seed words were set based on descriptions of nine types of climate events (heavy rain, hail, low temperature, drought, high temperature, lightning, sandstorm, typhoon, and smog) in the *China meteorological risk event yearbook*<sup>25</sup>. Expanding the semantic similarity of these climate-related seed words and conducting manual screening produced a climate feature dictionary containing 292 words<sup>26</sup>.

Secondly, the climate loss concern dictionary consisted of keywords that describe the public’s expressions of losses related to climate events. To develop a more accurate dictionary of climate loss concern, we used risk-related words and their synonyms defined by Loughran & McDonald (2011) as seed words<sup>27</sup>. The distances between the combined vectors of climate feature information and risk concern information were calculated based on semantic similarity in the vector space. Words exhibiting proximity to both categories were subsequently selected. This resulted in a climate loss concern dictionary containing 1,820 words.

Specifically, combined vectors for climate features and loss concern information were extracted using the approach described above. The climate feature information was sourced from the preconstructed climate feature dictionary, while the risk concern information for climate events was extracted by combining the cosine similarity of climate and risk word vectors. The sentiment dictionary encompassed two parts. To capture the public’s

concerns regarding climate events, mainstream sentiment dictionaries were referenced, including those from Baidu<sup>28</sup>, the Harbin Institute of Technology<sup>29</sup>, Jiang *et al.*<sup>30</sup>, and Ma *et al.*<sup>31</sup>, and aggregated to form two sentiment dictionaries, one for positive emotions and another for negative emotions.

**Data preprocessing.** The *Message Board for Leaders* underwent two major updates (the first in 2020 and the second in 2022), so data from its three developmental stages were merged. The merged fields included the link, user nickname, message title, message time, response time, and response content. Further text cleaning was conducted to ensure data quality and improve the accuracy and reliability of the model. The cleaning process primarily involved removing irrelevant information, such as punctuation, URLs, and extra spaces. For text segmentation, the pkuseg toolkit was used, which offers higher accuracy than the commonly used jieba toolkit<sup>32</sup>.

The messages were then identified using the climate feature dictionary, resulting in the extraction of messages related to nine main types of climate events in China: heavy rain and flooding, hail and tornadoes, low temperature and freezing damage, drought, heatwaves, lightning, sandstorms, typhoons, and smog. These messages, totaling 222,923 entries, formed the text dataset of expressions of public views on climate risk.

**Calculation of the expressions of public views index on climate risk.** The expressions of public views index on climate risk was constructed based on descriptions of potential losses contained in each regional expression. First, the expressions of public views situation index  $\text{expressions}_{n,t}$  for each text was calculated as follows:

$$\text{opinion}_{n,t} = 100 \times \left( \frac{RW_{n,t}}{TW_{n,t}} \right) \times \left( \frac{NW_{n,t} - PW_{n,t}}{NW_{n,t} + PW_{n,t}} + 1 \right) / 2 \quad (1)$$

where  $RW_{n,t}$  represents the number of loss concern words in the  $n$ -th message on day  $t$ , sourced from the climate loss concern dictionary.  $TW_{n,t}$  represents the total number of words in the  $n$ -th message on day  $t$ .  $PW_{n,t}$  represents the number of positive words in the  $n$ -th message on day  $t$ , sourced from the positive sentiment dictionary.  $NW_{n,t}$  represents the number of negative words in the  $n$ -th message on day  $t$ , sourced from the negative sentiment dictionary.  $\frac{RW_{n,t}}{N_{n,t}}$  represents the percentage of loss concern words in the message, which accounts for the variability in message length and provides a measure of the public's loss concern.  $\left( \frac{NW_{n,t} - PW_{n,t}}{NW_{n,t} + PW_{n,t}} + 1 \right) / 2$  measures the sentiment of the message, ranging from 0 (most positive) to 1 (most negative). Thus, the score of the expressions of public views at the message level can be interpreted as the loss measure weighted by the sentiment of the message. When the message is more negative (or positive), it has a higher (or lower) weight, reflecting the impact of sentiment on the perceived loss.

To capture the degree of expressions of public views on climate risk in a region, the daily expressions score  $\text{day\_expressions}_t$  was calculated by aggregating the message-level expressions scores:

$$\text{day\_opinion}_t = \sum_{n=1}^{N_t} \text{opinion}_{n,t} \quad (2)$$

where  $N_t$  represents the total number of messages about climate risk expressions of public views posted on day  $t$ .

Next, the daily indices were aggregated into monthly and annual indices to provide a broader temporal perspective. The monthly and annual expressions of public views indices for climate risk are calculated as follows:

$$\text{month\_opinion}_m = \sum_{t \in T_m} \text{day\_opinion}_t \quad (3)$$

$$\text{year\_opinion}_y = \sum_{t \in T_y} \text{day\_opinion}_t \quad (4)$$

where  $T_m$  represents the set of all days in month  $m$ .  $T_y$  represents the set of all days in year  $y$ ,  $t \in T_m$  indicates that  $t$  is a day in month  $m$ , and  $t \in T$  indicates that  $t$  is a day in year  $y$ .

**Calculation of the government responses index on climate risk.** The government responses index on climate risk was evaluated based on three dimensions: status, timeliness, and quality. The government responses for a given month, denoted as  $\text{month\_response}_m$ , was calculated by averaging the max-min standardized values of the response status index ( $\text{month\_response\_status}_m$ ), response timeliness index ( $\text{month\_response\_time}_m$ ), and response quality index ( $\text{month\_response\_quality}_m$ ):

$$\begin{aligned} \text{month\_response}_m = & (\text{month\_response\_status}_m + \\ & \text{month\_response\_time}_m + \\ & \text{month\_response\_quality}_m) / 3 \end{aligned} \quad (5)$$

Prior to averaging, the response status and response quality index were normalized using the *max-min* normalization method, whereas the response timeliness index was transformed using *min-max* normalization method to account for its inverse relationship with performance.

The  $month\_response\_status_m$  measures the proportion of government responses to climate-related expressions in a given period. It was calculated as follows:

$$month\_response\_status_m = \frac{a_m}{q_m} \quad (6)$$

where  $a_m$  is the number of responses in month  $m$  and  $q_m$  is the number of expressions in month  $m$ . The  $month\_response\_time_m$  measures the time taken for the government to respond to each message, reflecting the speed at which the government addresses expressions of public views and requests. It was calculated as follows:

$$response\_time_{m,n} = t_{m,n}^{reply} - t_{m,n}^{post} \quad (7)$$

where  $t_n^{reply}$  represents the response time for the expression  $n$  in month  $m$ .  $t_n^{post}$  represents the posting time for the expression  $n$  in month  $m$ . This metric reflects the efficiency of the government in addressing and responding to issues. The response times for all messages in a given region were then aggregated to obtain the monthly response timeliness index:

$$month\_response\_time_m = \frac{\sum_{n=1}^{N_m} response\_time_{m,n}}{N_m} \quad (8)$$

where  $N_m$  is the total number of messages in month  $m$ .

The *quality* dimension measures the relevance of the government's responses, reflecting whether the government directly addresses the issues raised by the public and provides reasonable responses<sup>33–35</sup>. Textual similarity can reflect the consistency of cognitive component matching, with high similarity indicating that the recipient has accurately captured the public's concerns<sup>36</sup>. This was measured using the soft cosine similarity between the message and the response, denoted as  $response\_quality_{m,n}$ , which represents the response relevance for the  $n$ -th message in month  $m$ . Soft cosine similarity takes into account semantic relationships, word weights, and variations and synonyms, thereby improving the accuracy and reliability of the similarity measure.

First, all words were vectorized by the Word2Vec model, and the relatedness between words was calculated to form the word relatedness matrix  $V$ . The element  $v_{i,j}$  in  $V$  is given by:

$$v_{i,j} = \max(0, \text{cosimilarity}(v_i, v_j)) \quad (9)$$

where  $v_i$  is the 300-dimensional word vector for word  $w_i$ ,  $v_j$  is the 300-dimensional word vector for word  $w_j$ .

The soft cosine similarity between the message and the response was then calculated as follows:

$$response\_quality_{m,n} = \frac{Q^T V A}{\sqrt{Q^T V Q} \times \sqrt{A^T V A}} \\ = \frac{\sum_{i=1}^n \sum_{j=1}^n q_{m,n,i} v_{i,j} a_{m,n,j}}{\sqrt{\sum_{i=1}^n \sum_{j=1}^n q_{m,n,i} v_{i,j} q_{m,n,j}} \times \sqrt{\sum_{i=1}^n \sum_{j=1}^n a_{m,n,i} v_{i,j} a_{m,n,j}}} \quad (10)$$

where  $q_{m,n,i}$  represents the word  $i$  in the expression  $n$  in month  $m$ , and  $a_{m,n,j}$  represents the word  $j$  in the corresponding response in month  $m$ . This index ranges from 0 (worst) to 1 (best), with no response automatically recorded as a null value.

The response quality for all messages in a given region was then aggregated to obtain the monthly response quality index:

$$month\_response\_quality_m = \frac{\sum_{n=1}^{N_m} response\_quality_{m,n}}{N_m} \quad (11)$$

where  $N_m$  is the total number of messages in month  $m$ .

## Data Records

The datasets for the expressions of public views index on climate risk and the government responses index on climate risk have been uploaded in Figshare<sup>37–39</sup>. Detailed index information is provided in Table 1.

The national-level dataset<sup>37</sup> includes the expressions of public views index on climate risk and government responses index on climate risk for China from 2011 to 2023, with daily, monthly, and annual indicators. For expressions of public views on climate risk, sub-tables for common risk events, such as heavy rain, high temperature, and low temperature, are also provided. The original annual data for response status, response timeliness, and response quality are included.

The provincial-level dataset<sup>38</sup> includes the expressions of public views index on climate risk and government responses on climate risk for 31 provinces in China (excluding data from Hong Kong, Macau, and Taiwan) from 2011 to 2023, with monthly and annual indices. The city-level dataset<sup>39</sup> includes the expressions of public views index on climate risk and government responses index on climate risk for 293 cities in China (excluding Hong Kong, Macau, and Taiwan) from 2011 to 2023, with monthly and annual data.

| Level             | Content                                                                                                          | Filename                                                                                                           | Sheet Name        |
|-------------------|------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|-------------------|
| National          | National expressions of public views index on climate risk                                                       | China's expressions of public views index                                                                          | Monthly<br>Annual |
|                   | Expressions of public views Index on heavy rain, high temperature, low temperature, and drought                  | China's expressions of public views index on heavy rain, high temperature, low temperature, and drought            | Monthly<br>Annual |
|                   | National government responses index on climate risk                                                              | China's government responses index                                                                                 | Monthly<br>Annual |
| Provincial        | Expressions of public views index on climate risk for 31 provinces                                               | China's provincial expressions of public views index                                                               | Monthly<br>Annual |
|                   | Expressions of public views index on heavy rain, high temperature, low temperature, and drought for 31 Provinces | China's provincial expressions of public views index on heavy rain, high temperature, low temperature, and drought | Monthly<br>Annual |
|                   | Government responses index on climate risk for 31 provinces in China                                             | China's provincial government responses index                                                                      | Monthly<br>Annual |
| City              | Expressions of public views index on climate risk for 293 cities in China                                        | China's city-level expressions of public views index                                                               | Monthly<br>Annual |
|                   | Government responses index on climate risk for 293 cities in China                                               | China's city-level government responses index                                                                      | Monthly<br>Annual |
| Other Information | Original text data of expressions and responses                                                                  | Expressions and response dataset                                                                                   | /                 |
|                   | Climate feature dictionary and climate loss concern dictionary                                                   | Climate feature dictionary and climate loss concern dictionary in Chinese and English                              | /                 |

**Table 1.** Summary of the dataset.

A dataset of 222,923 paired expressions and responses is stored on Figshare<sup>40</sup>. Climate feature dictionary and climate loss concern dictionary are also stored on Figshare<sup>26</sup>.

### Technical Validation

The technical validation of this study encompassed three main parts: comparison with manual verification, comparison with existing datasets, and limitations. In the supplementary files, there is a related analysis of the index, including analyzing spatiotemporal patterns in the expressions of public views index on climate risk and examining the geographical distribution and correlation characteristics of the government responses index. Additionally, the supplementary materials also contain explanations of the background information features of the *Message Board for Leaders*. More detailed explanatory information about the response \_quality index has also been included in the supplementary files.

**Comparison with manual verification.** Manual verification primarily focuses on the validity of the word sets, including the robustness of climate feature identification and climate loss concern information. A manual audit team composed of master's and doctoral students in economics and finance was formed to determine whether a message contained climate features and climate risk concerns.

The robustness validation for climate feature identification focused mainly on the climate feature dictionary. A comparison experiment between human-labeled and machine-labeled data was conducted. Specifically, 10,000 randomly selected messages were used, of which 5,000 were climate loss-related messages (positive test samples) and 5,000 were non-climate loss-related messages (negative test samples). These messages were manually coded to identify differences between human-labeled and machine-identified messages. Three coders independently coded the same texts, and the average Cohen's kappa for the coding categories was 94.75%, indicating a high level of inter-coder agreement. Discrepancies among human readers were resolved by simple majority voting.

The human labels were then compared with the original machine-generated labels, and standard precision and recall metrics were calculated separately for the positive and negative test samples. The machine-identified results showed high consistency with human readers. The comparison results for each coder are shown in Table 2. The average accuracy was 89.73%, the average precision was 95.29%, the average recall was 83.61%, and the average F1 score was 89.05%. Therefore, it was concluded that the climate features identified by human readers and the machine method were largely consistent. Table 2 provides a detailed comparison of the performance of the three coders and the climate feature dictionary predictions on the test dataset, demonstrating the high consistency and reliability of the machine-identified climate features.

The robustness validation for climate loss concern information primarily focused on the climate loss concern dictionary. This study examined the alignment between machine-labeled discussions of climate loss and human readers' intuitive labeling of those discussions. Following Wu's sample selection approach, the combined test sample was composed of 900 randomly selected sentences<sup>41</sup>. First, 600 messages identified as climate loss-related (positive test samples) were randomly selected. Then, one instance of a climate loss discussion (i.e., any word close to the loss words in the supplementary vocabulary) was identified from each message, and the entire sentence surrounding this instance was extracted. Next, 300 non-climate loss messages (negative test samples) were randomly selected, and a sentence containing only a word from the loss vocabulary was extracted from each message.

|           | Experimenter 1 | Experimenter 2 | Experimenter 3 | Average |
|-----------|----------------|----------------|----------------|---------|
| Accuracy  | 89.16%         | 89.78%         | 90.24%         | 89.73%  |
| Precision | 96.57%         | 94.30%         | 95.01%         | 95.29%  |
| Recall    | 81.20%         | 84.68%         | 84.94%         | 83.61%  |
| F1 Score  | 88.22%         | 89.23%         | 89.69%         | 89.05%  |

**Table 2.** Performance Comparison of Three Coders' Verification Results and Climate Feature Dictionary Predictions on the Test Dataset.

Three coders were recruited to independently label each of the 900 sentences based on whether they contained any climate loss-related discussions. Standard precision and recall metrics were calculated using both the positive and the negative test samples. The machine-identified results showed high consistency with human readers, with an accuracy of 92.01%, precision of 99.78%, recall of 90.53%, and F1 score of 94.93%. Therefore, the automated method was considered highly consistent with the traditional climate loss concepts held by human readers and capable of accurately identifying climate loss-related discussions.

**Comparison with existing datasets.** Due to the lack of publicly accessible datasets that simultaneously cover both expressions of public views on climate risk and government responses, a direct comparison with existing datasets is not feasible. This study is one of the earliest to use message data to analyze public perceptions of climate risk and government responses. Existing datasets primarily focus on either public perceptions or government actions. Table 3 summarizes these datasets, detailing their scope and coverage years.

One dataset related to government actions is from Song *et al.*<sup>42</sup>, who developed an assessment indicator system for China's progress in adapting to climate change. This system evaluates the government's climate risk adaptation capabilities from dimensions such as monitoring and early warning, disaster reduction strategies, and infrastructure construction. Existing datasets on public perceptions often rely on Baidu Index, such as the one developed by Chen *et al.*<sup>43</sup> which focuses on public perceptions of macro-level climate change, and the dataset by Li & Tian<sup>44</sup> which examines the public perceptions of the direct impacts of climate change. Given their thematic similarity, we chose these datasets for comparison.

Firstly, in terms of data structure completeness and depth, our study is based on message data, where each record typically includes detailed information such as time, location, topic, and content, and has been reviewed and categorized by the platform. Compared to the Baidu Index used by Chen *et al.*<sup>43</sup> and Li & Tian<sup>44</sup>, our dataset provides detailed information for each message, allowing for a structured analysis of public expressions and government responses. Second, we emphasize the interaction between public expressions and government responses in the context of climate risk. By using machine learning and text analysis techniques to construct the index system, we can more accurately capture the dynamic relationship, rather than just performing aggregate statistical analyses as done by Song *et al.*<sup>42</sup>. In terms of spatiotemporal resolution, our dataset provides information at both monthly and annual time scales, covering national, provincial, and city levels. This is crucial for understanding the differences in attitudes toward climate change across different regions and how these attitudes evolve over time.

Additionally, we compare our dataset with two climate risk datasets that provide information at the city-daily level. Guo *et al.*<sup>45</sup> focuses on the characterization of physical climate risks, while Ma *et al.*<sup>31</sup> conducts further analysis of climate policy uncertainty. Distinct from these studies, our work specifically focuses on public awareness of climate risk. This complements existing studies on physical climate risks and policy uncertainty, contributing to a more comprehensive framework for understanding climate risk.

**Further limitations.** First, in terms of data traceability, the anonymous nature of message board data makes it difficult to obtain relevant users' socio-demographic information. Therefore, when social media data is used to extract the public's expressions of climate risks, the dispersiveness and unstructured nature of data sources are inevitably faced.

Furthermore, the heterogeneity of the data sources restricts the scope of this investigation. While the *Message Board for Leaders* is an important platform for collecting public expressions, it is not the only channel. Other channels, such as the Mayor's Hotline, may also gather climate-related expressions of public views. However, the information from these hotlines is typically non-public, and researchers cannot directly access or utilize these data sources. This, to some extent, affects the comprehensiveness of the study.

## Usage Notes

This dataset provides a comprehensive framework for studying Chinese public attitudes toward climate risk and government responses, with broad applicability in research on climate governance, public policy, and risk communication.

First, the dataset can be used to evaluate the effectiveness of local climate policies and improvements in disaster management capacity. By analyzing the government responses, researchers can assess how well local governments address public concerns and mitigate climate risks and compare their performance before and after policy implementation to identify best practices and areas for improvement. For instance, researchers can examine the spatiotemporal dynamics of both the expressions of public views index on climate risk and the government responses index on climate risk in specific regions, integrating external variables such as frequency of

| Dataset Name                                                                           | Level      | Granularity | Content                                                                                                                                                                                                                                     | Year Coverage | Source                           |
|----------------------------------------------------------------------------------------|------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|----------------------------------|
| Climate Risk Expressions of Public Views Index and Government Responses Index in China | City       | Monthly     | Expressions of public views and government responses on climate risk based on data from the <i>Message Board for Leaders</i>                                                                                                                | 2011–2023     | Our Data                         |
| China's Climate Change Adaptation Progress Assessment Indicators                       | National   | Annual      | Comprehensive evaluation system with 45 indicators, including meteorological observations and public health, based on data from the <i>National Development and Reform Commission</i> and the <i>National Bureau of Statistics of China</i> | 2010–2022     | Song <i>et al.</i> <sup>42</sup> |
| China's Climate Change Attention Index                                                 | National   | Monthly     | Baidu Search Index (keywords related to macro-level climate change)                                                                                                                                                                         | 2011–2020     | Chen <i>et al.</i> <sup>43</sup> |
| China's Climate Risk Perception Index                                                  | Provincial | Annual      | Baidu Search Index (keywords related to direct impacts of climate change, such as central weather station, train suspension)                                                                                                                | 2005–2022     | Li & Tian <sup>44</sup>          |
| China's Physical Climate Risk Dataset                                                  | City       | Daily       | Comprehensive physical climate risk index derived from daily observational data from <i>China Meteorological Administration</i>                                                                                                             | 1993–2023     | Guo <i>et al.</i> <sup>45</sup>  |
| China's Climate Policy Uncertainty Index                                               | City       | Daily       | Climate policy uncertainty index based on news reports from six major newspapers                                                                                                                                                            | 2000–2022     | Ma <i>et al.</i> <sup>18</sup>   |

**Table 3.** The Existing Dataset Overview.

extreme weather events and economic losses, to evaluate the effectiveness of initiatives like *Sponge City*, *Resilient City*, and early warning systems for extreme weather.

Additionally, the dataset supports assessments of disaster management outcomes in specific regions. By linking the expressions of public views index on climate risk with the government response index on climate risk, it reveals regional preparedness and responsiveness to climate-related disasters, as well as whether public concerns are effectively addressed. Firstly, researchers can combine physical climate risk indicators<sup>45</sup> to select regions with similar climatic conditions but differing governance outcomes, and compare the alignment between public concerns and governmental responses. Regions with high alignment can then serve as case studies to summarize successful practices in climate risk interactive governance. Secondly, time-series analyses can track changes in public risk expressions before and after specific disaster events (e.g., flood or heatwave), in conjunction with the timeliness and quality of government responses, to evaluate the actual effectiveness of disaster management efforts.

Moreover, the dataset enables comparative studies of regional differences in the risk expressions of public views on climate risk, helping to understand how varying climatic conditions and disaster experiences shape public risk awareness and concern levels. Researchers can use spatial econometric models to analyze the spatial distribution patterns of public climate risk expressions, exploring differences and underlying drivers between geographic regions, thus providing evidence-based support for local governments in formulating targeted climate adaptation policies.

In summary, this dataset can be used to evaluate the effectiveness of climate risk governance, regional climate adaptation, and resilience-building initiatives, and to inform policy recommendations aimed at improving climate governance and disaster management.

### Code availability

The codes that were used to generate the climate risk expressions of public views and government responses indices are available in Figshare<sup>46</sup>.

Received: 2 January 2025; Accepted: 14 May 2025;

Published online: 27 May 2025

### References

- Kreibich, H. *et al.* The challenge of unprecedented floods and droughts in risk management. *Nature* **608**, 80–86 (2022).
- Moussaïd, M., Brighton, H. & Gaissmaier, W. The amplification of risk in experimental diffusion chains. *Proc. Natl. Acad. Sci.* **112**, 5631–5636 (2015).
- Drews, S. & Bergh, J. C. J. M. van den. What explains public support for climate policies? A review of empirical and experimental studies. *Clim. Policy* **16**, 855–876 (2016).
- Petzold, J. *et al.* A global assessment of actors and their roles in climate change adaptation. *Nat. Clim. Change* **13**, 1250–1257 (2023).
- Anderson, C. C., Moure, M., Demski, C. & Renaud, F. G. Risk tolerance as a complementary concept to risk perception of natural hazards: A conceptual review and application. *Risk Anal.* **44**, 304–321 (2024).
- Bostrom, A., Hayes, A. L. & Crosman, K. M. Efficacy, action, and support for reducing climate change risks. *Risk Anal.* **39**, 805–828 (2019).
- Eisenack, K. *et al.* Explaining and overcoming barriers to climate change adaptation. *Nat. Clim. Change* **4**, 867–872 (2014).
- Kenny, J., Geese, L., Jordan, A. & Lorenzoni, I. A framework for classifying climate change questions used in public opinion surveys. *Environ. Polit.* 1–27 <https://doi.org/10.1080/09644016.2024.2429264> (2024).
- Wu, M. *et al.* Spatio-temporal difference analysis in climate change topics and sentiment orientation: Based on LDA and BiLSTM model. *Resour. Conserv. Recycl.* **188**, 106697 (2023).
- Wei, X., Zhang, Z., Zhang, M., Chen, W. & Zeng, D. D. Combining crowd and machine intelligence to detect false news on social media. *Manag. Inf. Syst. Q.* **46**, 977–1008 (2022).
- Wang, D., Zhou, Y., Qian, Y. & Liu, Y. The echo chamber effect of rumor rebuttal behavior of users in the early stage of COVID-19 epidemic in China. *Comput. Hum. Behav.* **128**, 107088 (2022).
- Lottering, S. J., Mafongoya, P. & Lottering, R. T. Assessing the social vulnerability of small-scale farmer's to drought in uMsinga, KwaZulu-Natal. *Int. J. Disaster Risk Reduct.* **65**, 102568 (2021).
- Murfin, J. & Spiegel, M. Is the risk of sea level rise capitalized in residential real estate? *Rev. Financ. Stud.* **33**, 1217–1255 (2020).
- Pankratz, N., Bauer, R. & Derwall, J. Climate change, firm performance, and investor surprises. *Manag. Sci.* **69**, 7352–7398 (2023).
- Allan, J. N. *et al.* Geographic distributions of extreme weather risk perceptions in the United States. *Risk Anal.* **40**, 2498–2508 (2020).

16. Faccini, R., Matin, R. & Skiadopoulos, G. Dissecting climate risks: Are they reflected in stock prices? *J. Bank. Finance* **155**, 106948 (2023).
17. Nagar, V. & Schoenfeld, J. Measuring weather exposure with annual reports. *Rev. Account. Stud.* **29**, 1–32 (2024).
18. Ma, Y.-R. *et al.* A news-based climate policy uncertainty index for China. *Sci. Data* **10**, 881 (2023).
19. Geese, L., Sullivan-Thomsett, C., Jordan, A. J., Kenny, J. & Lorenzoni, I. Measuring climate mitigation policy content in text-as-data: navigating the conceptual challenges. *Polit. Res. Exch.* **6**, 2387120 (2024).
20. Hu, D., Wang, E., Ye, Q., Chen, S. & Gu, X. How do online public messages affect local government responsiveness in China? A multilevel analysis based on the Message Board for Leaders. *J. Chin. Polit. Sci.* (2023).
21. Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W. & Starnini, M. The echo chamber effect on social media. *Proc. Natl. Acad. Sci.* **118**, e2023301118 (2021).
22. Yuan, Z. & Hu, W. Urban resilience to socioeconomic disruptions during the COVID-19 pandemic: Evidence from China. *Int. J. Disaster Risk Reduct.* **91**, 103670 (2023).
23. Lei, L., Zhang, D., Ji, Q., Guo, K. & Wu, F. A text-based managerial climate attention index of listed firms in China. *Finance Res. Lett.* **55**, 103911 (2023).
24. Ardia, D., Bluteau, K., Boudt, K. & Inghelbrecht, K. Climate change concerns and the performance of green vs. brown stocks. *Manag. Sci.* **69**, 7607–7632 (2023).
25. China Meteorological Administration. *China Meteorological Risk Event Yearbook in 2022*. (China Meteorological Press, Beijing, 2023).
26. Sun, X. *et al.* Climate feature dictionary and climate loss concern dictionary. *Figshare. Dataset*. <https://doi.org/10.6084/M9.FIGSHARE.28524848> (2025).
27. Loughran, T. & McDonald, B. Textual analysis in accounting and finance: A survey. *J. Account. Res.* **54**, 1187–1230 (2016).
28. Tian, H. *et al.* SKEP: Sentiment Knowledge Enhanced Pre-training for Sentiment Analysis. Preprint at <https://doi.org/10.48550/arXiv.2005.05635> (2020).
29. Tang, D., Wei, F., Qin, B., Zhou, M. & Liu, T. Building Large-Scale Twitter-Specific Sentiment Lexicon: A Representation Learning Approach. in *Proceedings of Coling 2014, the 25th International Conference on Computational Linguistics* 172–182 (2014).
30. Jiang, F., Lee, J., Martin, X. & Zhou, G. Manager sentiment and stock returns. *J. Financ. Econ.* **132**, 126–149 (2019).
31. Ma, F., Cao, J., Wang, Y., Vigne, S. A. & Dong, D. Dissecting climate change risk and financial market instability: Implications for ecological risk management. *Risk Anal.* 1–27 (2023).
32. Luo, R. *et al.* PKUSEG: A toolkit for multi-domain Chinese word segmentation. Preprint at <https://doi.org/10.48550/arXiv.1906.11455> (2022).
33. Lee, J. Can investors detect managers' lack of spontaneity? Adherence to predetermined scripts during earnings conference calls. *Account. Rev.* **91**, 229–250 (2016).
34. Cohen, L., Malloy, C. & Nguyen, Q. Lazy prices. *J. Finance* **75**, 1371–1415 (2020).
35. Hoberg, G. & Phillips, G. Text-based network industries and endogenous product differentiation. *J. Polit. Econ.* **124**, 1423–1465 (2016).
36. Te'eni, D. Review: A cognitive-affective model of organizational communication for designing IT. *Manag. Inf. Syst. Q.* **25**, 251 (2001).
37. Sun, X. *et al.* China's expressions of public views indices on climate risk and government responses on climate risk. *Figshare. Dataset*. <https://doi.org/10.6084/M9.FIGSHARE.28120541> (2025).
38. Sun, X. *et al.* China's provincial level expressions of public views indices on climate risk and government responses. *Figshare. Dataset*. <https://doi.org/10.6084/M9.FIGSHARE.28120544> (2025).
39. Sun, X. *et al.* China's city level expressions of public views indices on climate risk and government responses on climate risk. *Figshare. Dataset*. <https://doi.org/10.6084/M9.FIGSHARE.28120874> (2025).
40. Sun, X. *et al.* Expressions and responses dataset. *Figshare. Dataset*. <https://doi.org/10.6084/M9.FIGSHARE.28120850> (2025).
41. Wu, D. (Andrew). Text-based measure of supply chain risk exposure. *Manag. Sci.* **70**, 4781–4801 (2024).
42. Song, Y., Yang, X., Li, H. & Liu, M. Developing an indicator system and assessing China's progress on climate change adaptation in 2010–2022 from dual-dimension. *Environ. Sustain. Indic.* **26**, 100613 (2025).
43. Chen, X., Chen, X., Xu, L. & Wen, F. Attention to climate change and downside risk: Evidence from China. *Risk Anal.* **43**, 1011–1031 (2023).
44. Li, X. & Tian, Q. Haze pollution, climate risk perception and demand for commercial health insurance. *Sage Open* **14**, 21582440241242544 (2024).
45. Guo, K., Ji, Q. & Zhang, D. A dataset to measure global climate physical risk. *Data Brief* **54**, 110502 (2024).
46. Sun, X. *et al.* Codes used to generate the climate risk expressions of public views and government responses indices. *Figshare. Software*. <https://doi.org/10.6084/M9.FIGSHARE.28120904> (2025).

## Acknowledgements

This work was supported by National Social Science Fund of China (Grant No. 23&ZD093, 23AZD071), and National Science Fund of China (Grant No. 72348003, 72071197).

## Author contributions

X.S.: Research design, supervision, scientific discussion, manuscript revision, funding support. L.L.: Data processing, conducting experiments, drawing figures and tables, writing original manuscript. Y.-R.S.: Data processing, conducting experiments, results analysis, scientific discussion, manuscript revision. Y.-J.S.: Data collection, data processing and results analysis. D.Z. & Q.J.: Supervision, results analysis, scientific discussion, graph optimization, funding support.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-05207-1>.

**Correspondence** and requests for materials should be addressed to D.Z. or Q.J.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025