J Vis (2019) 22:197–213
https://doi.org/10.1007/s12650-018-0522-2

REGULAR PAPER

Bo Yang · Weiqun Cao · Chengming Tian

Online visual analysis of forest diseases

Received: 28 January 2018 / Revised: 25 April 2018 / Accepted: 26 April 2018 / Published online: 23 October 2018

Abstract It is of great practical significance to analyze and demonstrate more effectively the occurrence of forest diseases, which is an important subfield in the study of forest disasters. On the basis of such features of forest disease occurrence as timing, geography and disaster grade, we synthetically employ multiple visual elements and multiple interactive technologies to design and realize a forest disease visual analysis system to assist researchers and decision makers on issues related to forest diseases. With cases selected from the data set of forest diseases, we illustrate that the system is user-friendly and the applied visualization methods are effective.

Keywords Forest diseases · Visual analysis · Cluster analysis · Visualization

1 Introduction

Forests are widely recognized as the world's largest repository of land biodiversity and a significant factor in slowing global climate change. Forest disaster is an important cause for forest resource reduction in China, which is one of the most disaster-stricken countries in the world, with diverse disaster types, high percentage of disaster occurrence and wide area of disaster coverage. According to statistical data, from 1965 to 2015 China had roughly 700,000 cumulative forest disasters that affected almost 25% of China's forest areas. In average, there occurred 14,000 forest disasters of different levels each year. These numbers put China as the most disadvantageous victim of forest disasters in the world. Preliminary statistics shows that the direct economic losses caused by forest disasters equal approximately $150 million (Liu 2016). In recent years, massive forestry data have been digitized in the process of forestry informatization that raises urgent demand for new methods for data display and analysis. Forest disease is an important subfield in the study of forest disasters, which significantly affect the sustainability of forestry development. The typical methodology used for existing forest disease study is statistical analysis with simple display mode, which is unable to directly render the complex correlations among variables. The key task for this paper is to realize multi-dimensional analysis of forest diseases on the basis of both more effective analysis and demonstration of

B. Yang · W. Cao (✉)
School of Information Science and Technology, Beijing Forestry University, Beijing, China
E-mail: weiqun.cao@126.com

B. Yang
E-mail: yb2219@163.com

C. Tian
College of Forestry, Beijing Forestry University, Beijing, China
E-mail: chengmt@bjfu.edu.cn

forest disease occurrence, and sufficient integration of such disease features as timing, geography and disaster grade.

As an important tool to extract knowledge from large-scale complex data, visual analysis plays a key role in many fields. In forestry studies, however, the application of visual analysis is still in great need to employ its visual and interactive technologies to assist analysis of forest diseases.

On the basis of data analysis, model construction and framework design, we introduce a Web-based visual analysis system to demonstrate and analyze the occurrence of forest diseases. The system, which integrated visual analysis methods into forest disease data analysis, has great advantage on visual rendering over the traditional approach. The system visually displays the internal features of forest disease data from multiple dimensions to assist students and decision makers for more effective analysis. Specifically, our system makes the following contributions:

1. Display of forest disease occurrence in the temporal and spatial dimensions: our system enables users to learn the occurrence and trends of forest diseases through the interactions of multiple visual components to dynamically display the temporal spatial changes of these forest diseases.
2. Cluster analysis of forest disasters in the geographical dimension: our system assists users to learn the similarities and differences between forest diseases in different regions through cluster analysis in the geographical dimension.
3. Display of severity levels of occurring forest diseases: the system enables users to learn the severity of forest diseases through displaying diseases of different severity levels simultaneously in multiple views.
4. Solving the problem of partial display persistent to single view method: the system assists users to carry out interactive analysis for more effective mining of data through employing the method of multiple view comparison.

2 Literature review

2.1 The occurrence and trend of forest disasters

Climate change and pollution cause severe forest diseases in many regions. Some scholars analyze the features, distribution situations and trends of forest disasters by using the traditional field investigations and sample analysis methods and the function of Excel chart and then put forward relevant prediction models. Shang et al. (2015) analyze the geographic distribution of *Obolodiplosis robiniae*, which is an invasive insect pest, by sampling survey method. Hu et al. (2016) conduct a study on the occurrence and distribution of apple diseases in China and analyze the types and distributions and disease severity of apple diseases, by tabular recording. He and Lu (2018) use Excel to organize the data of the diseases and insect pests of *Pinus massoniana* Lamb. in Guizhou Province and analyze incidence, disaster grades, and the types and harms of the main diseases and pests. Li (2018) carries out the investigation and study on the types of fungal diseases in Guanmenshan Region in Benxi Manchu Autonomous County and records the results of the investigation and analysis of the disease pathogenesis. Wen et al. (1999) analyze the statuses and trends of forest diseases and pests in 1998 and 1999 in Jiangxi Province and discuss their characteristics and regularities. Hu et al. (2013) statistically analyze the distribution of pests and their harms in Chinese fir plantation in Kaihua county of Zhejiang Province and propose a series of control measures. Huang (2010) analyzes the cycles of major forest diseases in Sanming of Fujian Province from 1993 to 2007 with the superposition trend model and thereafter uses the gray catastrophe model to establish a prediction model. Some scholars use graphic image technology to analyze the characteristics of forest disease distributions. Peng et al. (2003) use microcomputer to automatically map such information of forest diseases as their distributions, types and disaster grades to provide a basis for engineering design for insect pest control and prevention. Zhang et al. (2014) analyze the occurrence area, occurrence degrees, occurrence rate, loss and year of Chinese diseases by multivariate linear regression method, combined line chart and bar chart. Chen et al. use geostatistical methods to analyze the spatial distribution characteristics of *Arhopalus rusticus* larvae and adults. They analyze the optimal fitting model of population spatial distribution through the curve graph of variation function and analyze the distribution characteristics of *Arhopalus rusticus* Linnaeus on Chinese pine by the vertical distribution chart and then analyze the intensive spatial aggregation of larvae and the randomness of adult in spatial distribution (Chen et al. 2018). Yang et al. (2018) collect and extract RNA from *Barley yellow striate mosaic virus* in Northern China and analyze the distribution and genetic diversity by image technology. Mayinuer et al. (2017) study species and distribution of cotton root stem diseases in eastern Xinjiang based on molecular detection. In addition, GIS is one of the commonly used analytical tools for forest disease research experts. Liu et al. (2016) use GIS to analyze the main disease distribution and disease index of alfalfa in Heilongjiang Province. Xu et al. (2018) conduct a survey and analysis on the species, distribution, hazard and occurrence characteristics of forest pests in Hongshan district of Chifeng city by GIS.

The literature review in this field shows three situations:

(1) The typical analytical method on forest disasters is statistical analysis whose results are verbally described;
(2) Static charts are typically used to display single dimension data while multi-dimensional information is largely neglected;
(3) GIS is commonly used while interactive visual analysis is absent.

2.2 Visualization of timing data

Forest diseases constantly vary with time and belong to typical sequence data. Time-oriented visualization is categorized into linear and periodic time visualizations according to the difference in the methods of presentation (Chen et al. 2013). Since the data of forest diseases used in this paper are linear, only the linear time series visualization method is discussed here. The time axis is one of the earliest and most widely used methods in linear time visualization (Aigner et al. 2011). Its horizontal axis represents the linear time domain, and the vertical axis represents the eigenvalue of the corresponding time domain. Typical time axis includes Gantt Chart (Wilson 2003), Lifeline (Plaisant et al. 2003; Wang et al. 2009), Storyline (Liu et al. 2013; Tanahashi and Ma 2012). The Gantt chart is usually used for project management to demonstrate the correlations between tasks and their arrangements more intuitively in graphics. The Lifeline visualizes the records of different patients with the time axis techniques. Storyline adopts river-based visual metaphors to show the effects of flow, merging, fork and disappearance in the time series. GapMinder converts international statistical data into lively, interactive and interesting charts to display the variance of multi-dimensional data over time with dynamic bubble diagrams (Rosling 2006).

2.3 Visualization of geographic information

Most studies of forest disasters on geographic information are based on GIS. Chen et al. (2014) use GIS to analyze the spatial distribution of forest fires in Chongqing by superimpose forest fires on the terrain. Li and Wang (2005) show the spatial distribution pattern of burned and affected areas of natural forests in Zhejiang Province with GIS. For geographic area data, a region of geographic space is a two-dimensional enclosed space with length and width identified by a series of points. Choropleth map (Torguson 2016) assumes that the data are evenly distributed within a region. Therefore, a specific feature such as population density and per capita income of each region can be represented with a single color. The spatial occurrence of forest diseases can also be displayed with Choropleth map.

2.4 Visualization of high-dimensional data

When parallel coordinates was first introduced by French mathematician d'Ocagne in 1885, it was limited to two-dimensional case, i.e., there are only two parallel axes in the plane (Ocagne 1885). Since Inselberg (1985) discusses parallel coordinates in multi-dimensional space and applies it in computational geometry, it gradually becomes one of the primary methods of high-dimensional data visualization and is widely used in various types of data visualization analysis systems. Parallel coordinates have such advantages as high space utilization rate, easy visual cognition and being highly interactive. A series of view transformation methods are introduced to reduce visual confounding of high-dimensional data. It is necessary to filter the data and reduce the data set to be visualized. The brush technology (Becker et al. 1987) is a visualization interactive technique, widely used in parallel coordinates, that can directly highlight the multi-dimensional data subset. Edge bundling techniques, mainly including force-directed edge bundling and hierarchical edge bundles technology, reduce visual complexity by using curves instead of straight lines and by reducing connections and crosses through clustered layout of wires (Holten 2006; Holten and Van Wijk 2009). Fisheye (Furnas 1986; Sarkar and Brown 1992) view transform is a focus information transform method that can gradually reduce context information around the focal point and therefore be able to visually highlight the key point while taking into account the peripheral information.

2.5 Visualization of hierarchical data

Treemap is a two-dimensional visualization method of space filling introduced by Johnson and Shneiderman (1991). It uses area and color to effectively demonstrate the distribution of leaf nodes in a tree structure. Its distinct advantage in efficient use of space makes itself an adaptive tool to display large-scale hierarchical dataset.

2.6 Clustering algorithm

Clustering is a type of unsupervised learning that puts similar objects into the same cluster. The commonly used clustering algorithm is the *k*-means clustering algorithm (Hartigan and Wong 1979) whose purpose is to divide n points into *k* clusters so that each point belongs to the cluster whose mean is closest to the point's value. This algorithm is used in cluster analysis module in our system.

3 Overview

This section first introduces the data used in this article, then explains the visual tasks of the system and finally introduces the pipeline of the visual analysis system.

3.1 Data

The statistical data of forest diseases in our system come from the series of *China Forestry Statistical Yearbook* (Bureau 2014). Based on the analytical goal of the system, this paper extracts and preprocesses data of year, region, disease occurrence and disease grade. The preprocessed data are shown in Table 1 involving more than 500 data records, each of which is graded in such three severity levels as mild, moderate and severe, in 31 provincial-level regions of China, excluding Hong Kong, Macao and Taiwan, over a time span of 17 years.

3.2 Analytical tasks

Through communications with experts on forest diseases, we summarize their key concerns into a list of analytical tasks.

T1. To survey the general situation of forest diseases at the national level.

- a. To show how forest diseases change over time at the national level.
- b. To compare the percentage changes of forest diseases that occur between adjacent years.
- c. To show different levels of forest disease occurrences.

T2. To survey forest diseases in current year.

- a. To display the geographical distribution of forest diseases at the national level.
- b. To compare the overall severity of forest diseases among different regions.

Table 1 Sample data of forest diseases

Year	Region	Occurrence area	Mild area	Moderate area	Severe area
1998	Beijing	800	650	150	0
1998	Tianjin	730	390	290	50
1998	Hebei	20,380	10,030	6600	3750
...	...	...	...	...	...
2014	Beijing	1975	1975	0	0
2014	Tianjin	6863	6148	582	133
2014	Hebei	27,884	25,076	2161	647
...	...	...	...	...	...

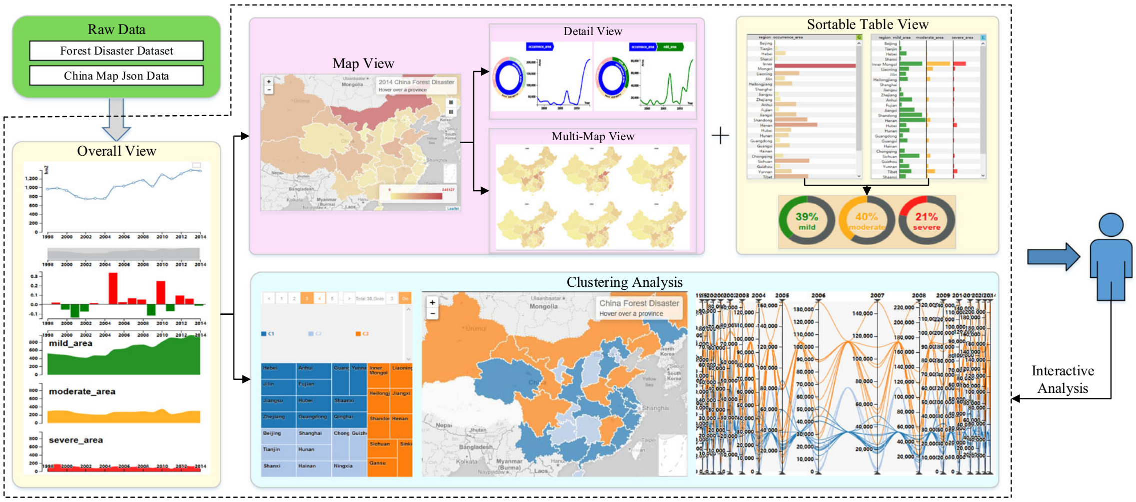

Fig. 1 System Overview. Raw Data: storage and processing of raw data; Overall View: visualizing the general survey of forest diseases over time; Map View: visualizing the spatial distribution of forest diseases; Sortable Table View: visualizing the severity comparison of forest diseases; Clustering Analysis: visualizing inter-regional similarities

  c. To compare the severity of forest diseases at all levels among different regions.

T3. To survey forest diseases in specific regions.
T4. To survey the overall situation of regional forest diseases.

  a. To dynamically display the geographical distribution of forest diseases in the country.
  b. To compare the regional evolutions of forest diseases in the country.

T5. To survey regional similarities of forest diseases.

3.3 Overview of the system

Our system employs information visualization and human–computer interaction technologies to analyze the temporal and spatial trends of forest diseases, and the inter-regional similarities and differences. Figure 1 shows the architecture of the system and the visual analysis pipelines. The system calls the relevant data and maps the structured data into graphic symbols, temporal variables, spatial variables, etc. through visual views. The system has four visual views of forest diseases including the overall view, map view, sortable table view and clustering analysis view. There are multiple interactive operations designed in the system including both intra-view and inter-view interactions.

4 Visual system

This section introduces the visual design of the system. The visual design methods for the overall view, map view, sortable table view and clustering analysis view are discussed.

4.1 The overall view

This view displays the annual changes of forest diseases with the line chart, diverging bar chart and multi-area chart. Different information is displayed through combinations of the three charts.

4.1.1 Line chart

The line chart shows the annual area changes of forest diseases (T1.A). The horizontal axis represents year while the vertical axis represents the occurrence area. A series of mouse interaction events are added to facilitate user analysis. Each hollow circle on the line chart represents the affected area in the current year in

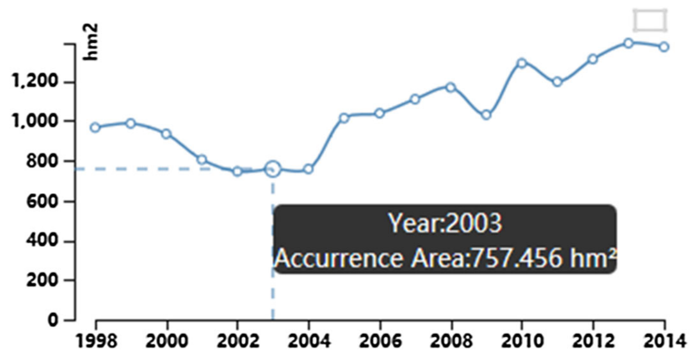

Fig. 2 Line Chart. The mouse points to the hollow circle of 2003 to display the corresponding information

the corresponding area. A hollow circle amplifies when the mouse hovers over to display horizontal and vertical dashed lines pointing to the axes with a box prompt to show the corresponding information as shown in Fig. 2. Mouse clicks on the hollow circles enable collaborate updates with the map views, the sortable table views and the parallel coordinates views. The scrolling of the mouse wheel zooms the yearly scope of display. When the yearly scope is small, left and right drags can be made with presses of the left mouse button on the blank area in the diagram to move back and forth in the current yearly scope size, as shown in Fig. 3. Brush functions can be turned on and off by clicking the rectangular selection button at the top right of the line chart to brush multi-year hollow circles to calculate the mean of affected areas of the

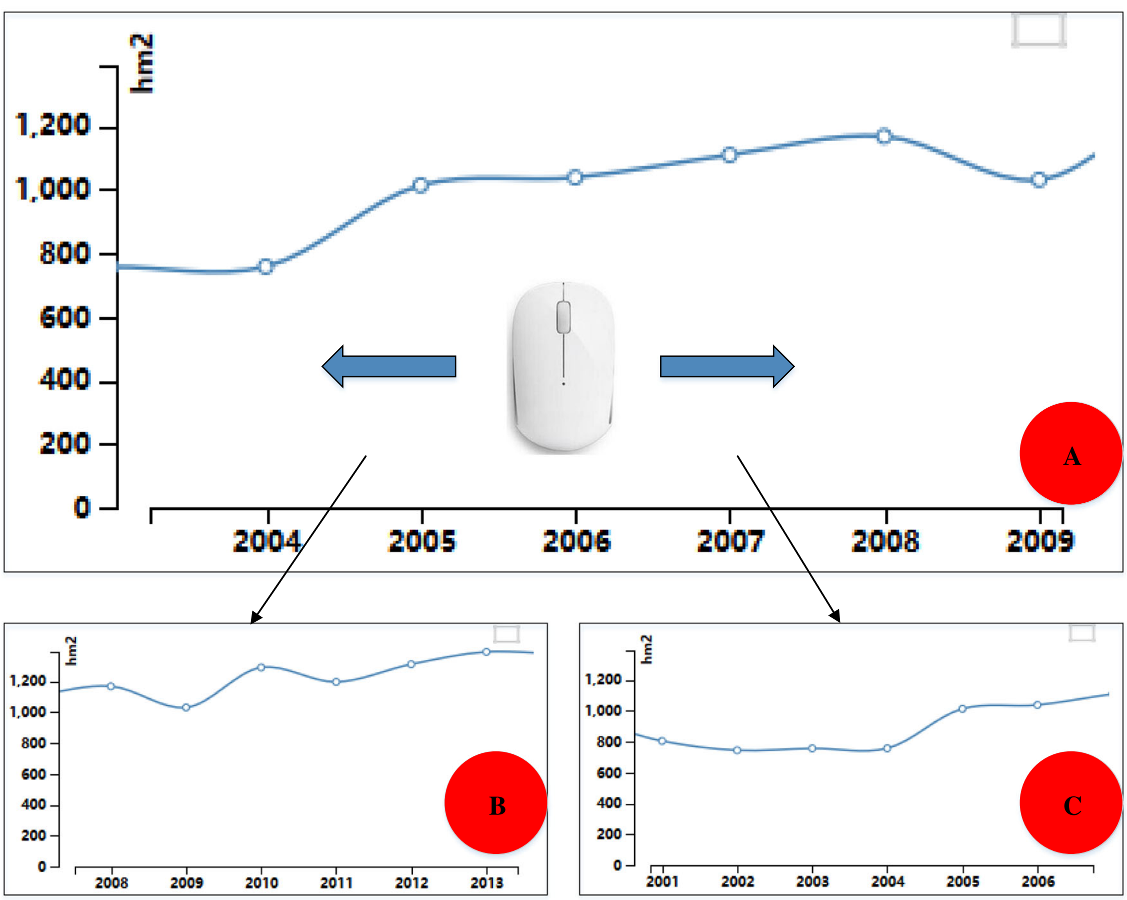

Fig. 3 (**A**) Line chart shows the data from 2004 to 2009. (**B**) Press the left mouse button to drag left in the chart, and the data from 2008 to 2013 are displayed. (**C**) Drag right in the line chart, and the data from 2001 to 2006 are displayed

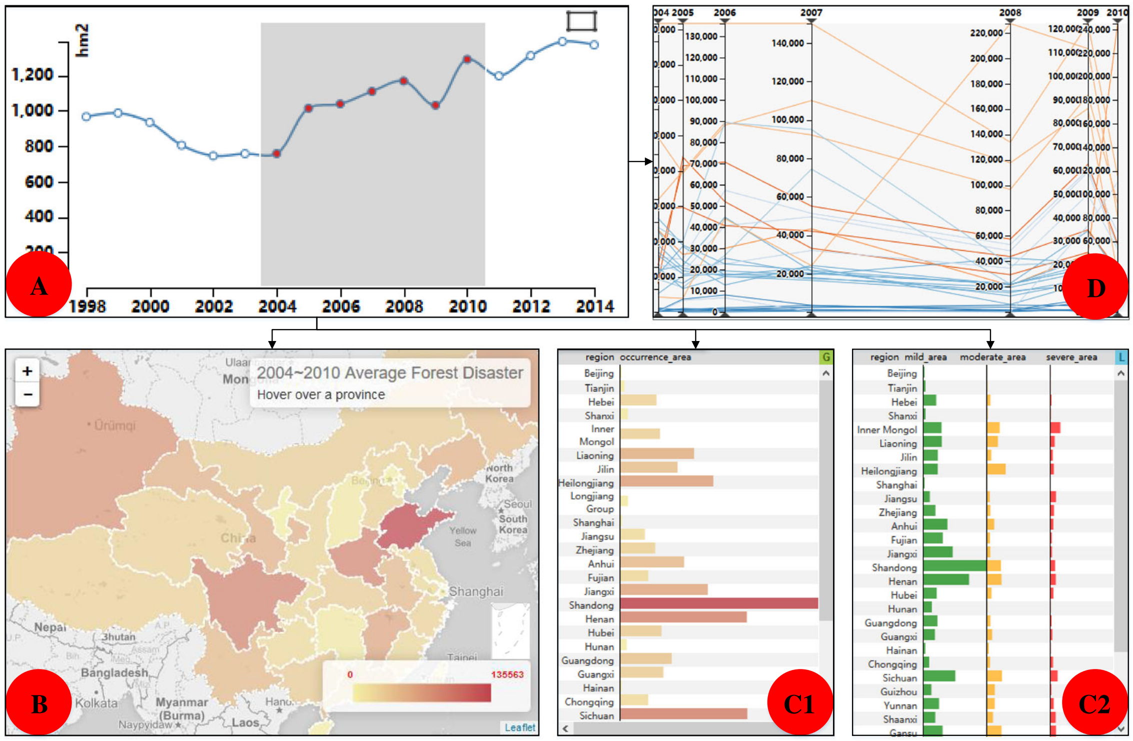

Fig. 4 When users choose data points 2004–2010 in (**A**), the average distributions of forest diseases in the selected scope are shown in (**B**). The overall ranking and the rankings of each disease grade of the regions are shown, respectively, in (**C1**) and (**C2**). The parallel coordinates display of all data points within the selected scope as shown in (**D**). Users can analyze the data through the collaborations of multiple diagrams

chosen years. Simultaneously, the map view, sortable table view and parallel coordinate view can realize collaborate updates, as shown in Fig. 4.

4.1.2 *Diverging bar chart*

Researchers need to compare the increase or decrease in the disease-affected areas between the current and previous years. Therefore, we use the diverging bar chart to show the annual percentage change of the affected areas (T1.b). The horizontal axis represents time, and the vertical axis represents percentage. A red rectangle above the horizontal axis represents a percentage increase, while a green rectangle below represents a percentage decrease. When the mouse is hovering over the rectangle, it is highlighted in orange, and the relevant information is displayed at the same time, as shown in Fig. 5. Other interactive events of the mouse are similar to those in the line chart. The equation to calculate percentage change is

$$P = (A_c - A_l)/A_l \quad (1)$$

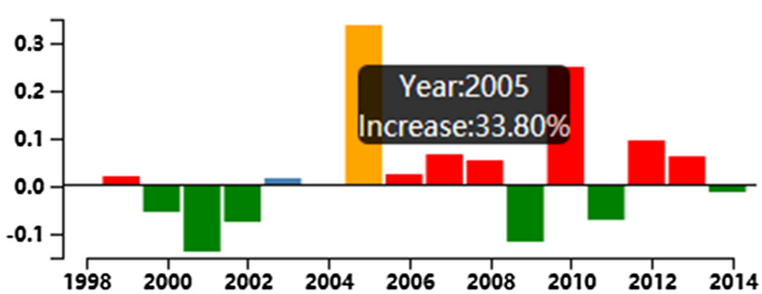

Fig. 5 Diverging Bar Chart. It shows a 33.8% increase from the previous year in 2005

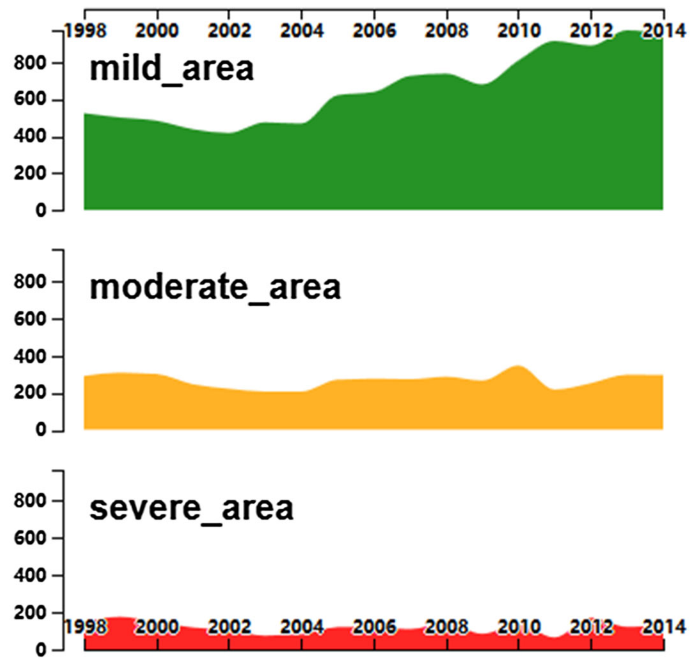

Fig. 6 Multi-area Chart. Forest diseases in various years are demonstrated in mild, moderate and severe levels

where A_c represents the affected area of the current year and A_l represents the affected area of the previous year. P is the percentage change.

4.1.3 Multi-area chart

The multi-area chart shows the annual changes of occurrence areas of three severity levels of forest diseases (T1.c). The horizontal axis represents year, and the vertical axis represents the occurrence area. Mild, moderate and severe are coded in green, orange and red, as shown in Fig. 6. When the mouse hovers over a specific year, a rose chart appears to display the percentage of the year's three levels, as shown in Fig. 7. The color coding corresponds to the multi-area chart, and the size of the area represents the percentage of the three degrees. The center area in the rose chart shows the current year.

4.1.4 Annual selector

The annual selector assists users to select years. The users can enlarge the scope of the annual display by operating the annual selector and realize the collaborations of the line chart, the diverging bar chart and the multi-area chart. Range transformation can be achieved through precise selection, horizontal size adjustment and move operation, as shown in Fig. 8. Through the overall view shown in Fig. 1, users can observe the

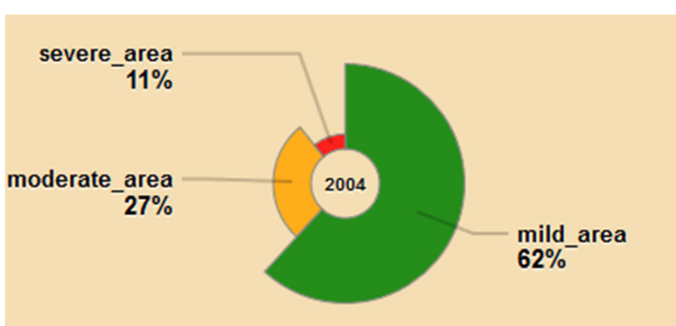

Fig. 7 Rose Chart. It shows the percentage of forest disease grades in 2004, with 62% mild, 27% moderate and 11% moderate

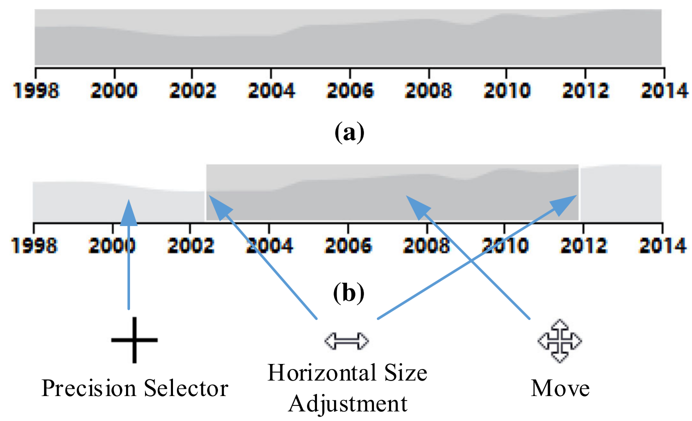

Fig. 8 Annual Selector. **a** The range of currently selected years: 1998–2014. **b** The range of currently selected years: 2003–2011. Precision Selection is applied to the blank area of the selector. Horizontal Size Adjustment is applied to the two ends of the slider. Move is operated on the slider

general changes of diseases. Various relevant information can be conveniently checked through interactions with the mouse.

4.2 Map view

The map view module is used to show the geographical distributions of the diseases. The analytical procedure for this module is shown in Fig. 9. The following sections detail the functions of each submodule.

4.2.1 Choropleth map

The Choropleth map is used to show the geographical distribution of the diseases in a given year (T2.a). As shown in Fig. 9, it displays the severity changes of disease occurrence areas with the gradient of colors. The

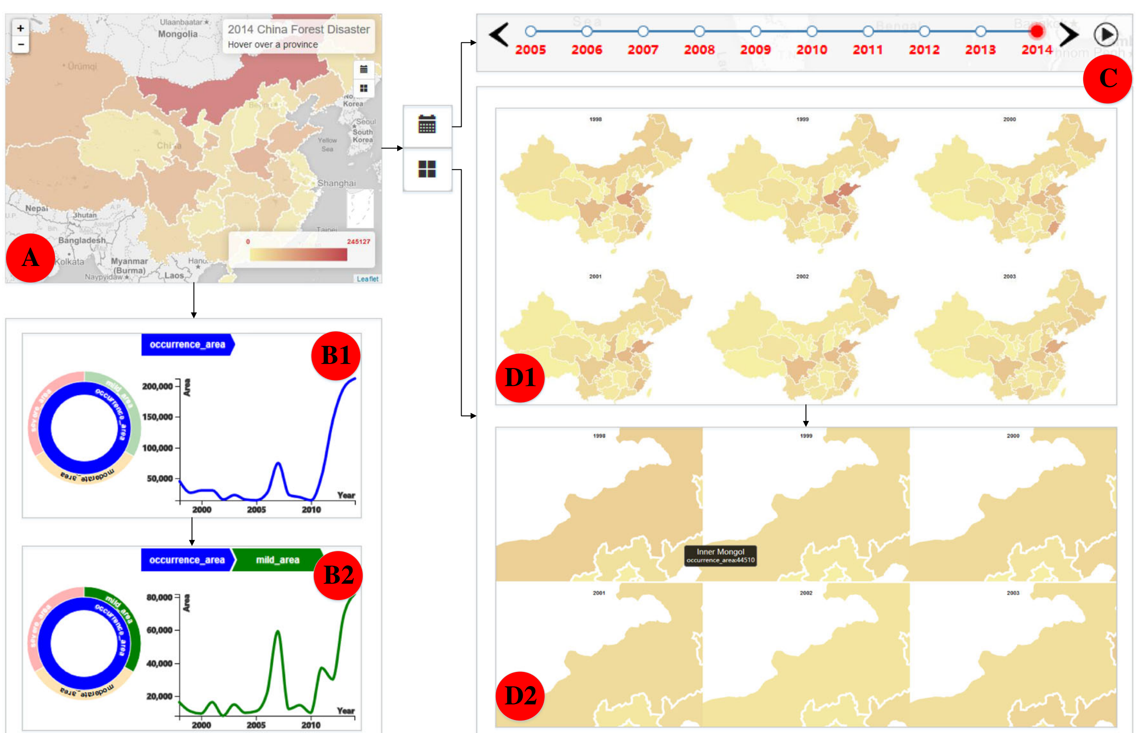

Fig. 9 Analytical flow chart of the Map Chart Module

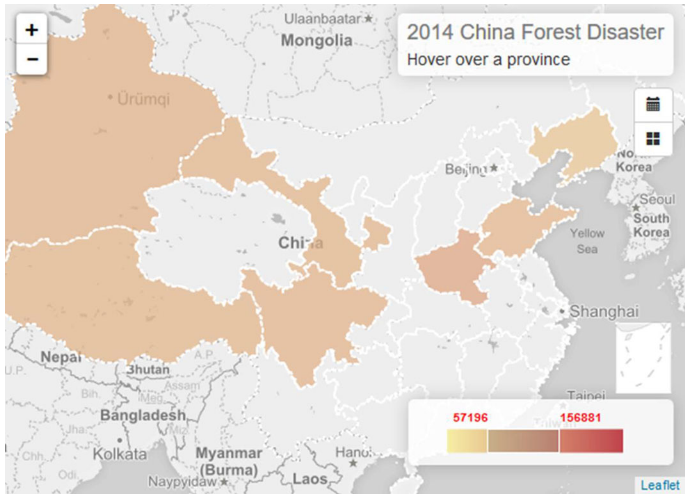

Fig. 10 Forest diseases in China in 2014 with a data arrangement from 57,196 to 156,881

lower right corner is area range selector, convenient for the user to check the distribution in a given region, as shown in Fig. 10. Detailed information about the selected area is displayed in the upper right corner when the mouse hovers over that area shows.

4.2.2 *Area details view*

A mouse click on a specific area in the map opens a popup window to display information of the corresponding region (T3). As shown in Fig. 9B1, the left ring indicates the total area and the selector for severity. The inner ring uses blue coding to represent the entire occurrence area while the outer ring coded in green, orange and red is to represent the mild, moderate and severe areas of disease occurrence, respectively. Meanwhile, opacity is set for the ring, which is highlighted when the mouse hovers, to increase user experience. When the mouse hovers the inner ring, the annual area changes of the corresponding region are shown in the right line chart, above which the indicator shows the currently selected option. As shown in Fig. 9B2, information of the mild area is displayed when the mouse hovers over the green ring.

4.2.3 *Dynamic display view*

A mouse click on the right calendar selector in the map pops up another view of the map where a time selector is added to the bottom left, as shown in Fig. 9C (T4.a). Users can click the start button on the time selector to automatically play the annual changes of the diseases. The play button switches into a pause button so that the animation can be paused in the user's convenience. A mouse click of a specific year on the time selector leads directly to the information of that year. The user can choose among the years by clicking the forward or backward buttons of the selector.

4.2.4 *Stamps view*

A mouse click on the right button in the map pops up a window to display the stamps view module, as shown in Fig. 9D1 (T4.b). Events occurred in every year are displayed in a single page for the users to observe and compare conveniently. A click on any area in any of the sub-maps can enlarge the area, as shown in Fig. 9D2, for the users to conveniently compare selected areas. The color code in this chart corresponds to the map. At the same time, an annual selector similar to that in Sect. 4.1.4 is added for the user to select a concerned area for display.

4.3 Sortable table view

The sortable table view is designed to assist users to compare the severity of diseases among different regions. This module ranks the severity of diseases in various regions in a given year. Figure 11A shows a sortable table on the overall situation of diseases in various regions of the current year (T2.b), the left column shows the name of the region, and the right column encodes the severity of diseases with rectangular whose length and color depth are positively correlated with the disease severity. Figure 11B shows a sortable table that shows the levels of diseases in the selected year (T2.c) with the left column showing the regional names while the right three columns showing the three severity levels of mild, moderate and severe coded in green, orange and red, respectively. Users can sort the data by clicking the title bars. Figure 11C, D

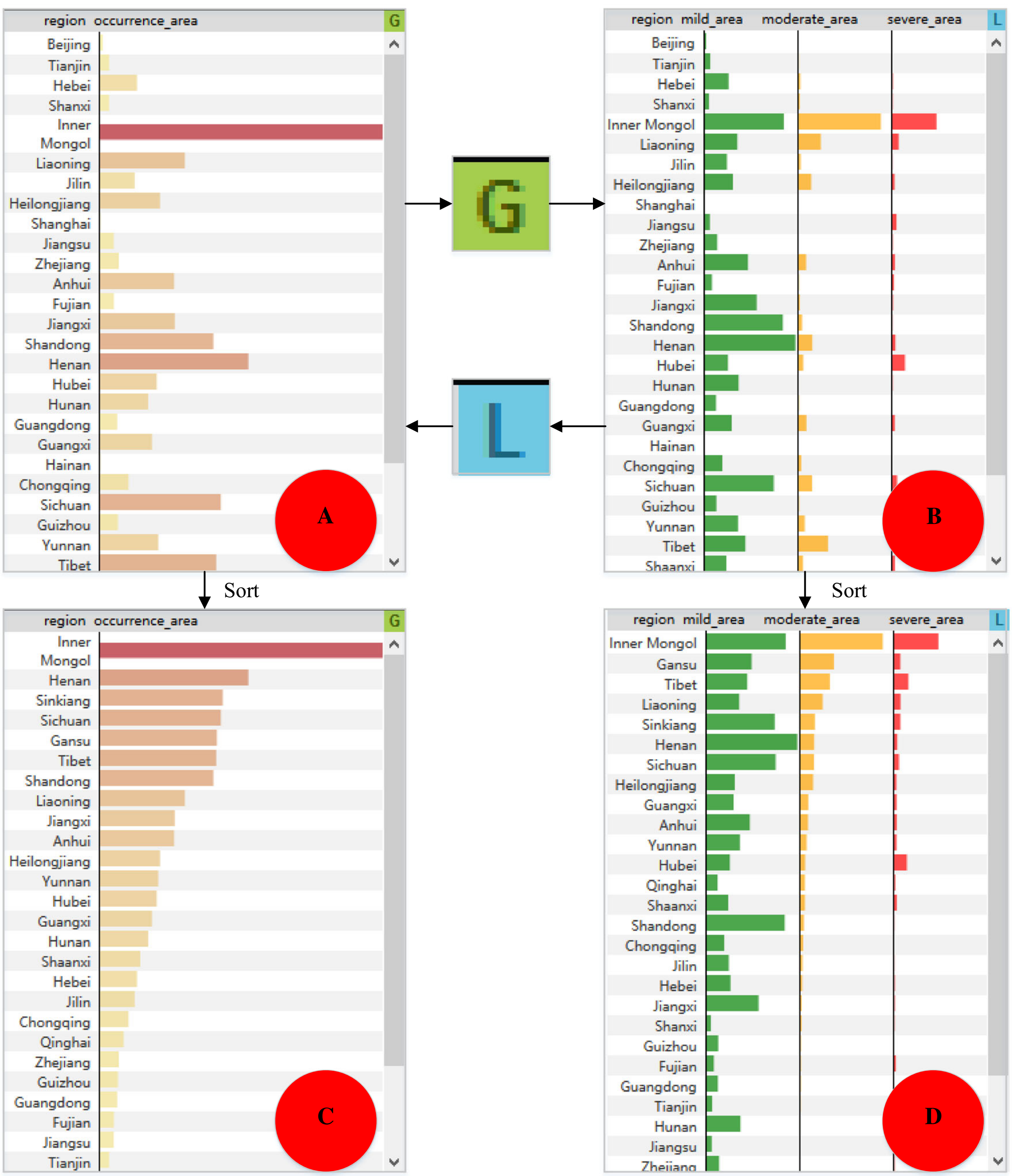

Fig. 11 Sortable table view

Fig. 12 Rings show the percentages of three forest disease levels

shows the sorted tables of 11A, B, respectively. Clicks on the icons in the upper right corner of the sorting table lead to switches between the overall situation and severity levels of the diseases. A mouse hovering over a row brings display of the percentages of disease severity in the corresponding area. As shown in Fig. 12, the percentages of the three levels are displayed in three rings whose central areas show the severity ratings and their percentages, respectively.

4.4 Cluster analysis view

This module assists users to discover the inter-regional similarities and differences in forest diseases (T5). The procedure of cluster analysis is shown in Fig. 13. *K*-means cluster algorithm is applied for regional cluster analysis, of which the cluster number can be selected with the cluster selector (Fig. 13A). At the same time, alternative input options are provided. Users can directly type the cluster number into the input box and then click the button Go to generate the corresponding visual effect directly. Total identifies the maximum cluster number, which is the number of regions that can be clustered. The forward and backward arrows assist users to directly generate the cluster numbers forward and backward. The system then generates the corresponding categorical identifications and legends according to the selected cluster number, as shown in Fig. 13B. The system generates the corresponding cluster results, which can be displayed with maps, parallel coordinates and treemap, respectively, as shown in Fig. 13C.

In the map module, cluster categories are shown in different cluster colors for users to conveniently observe the geographical distribution of clustered diseases. In the parallel coordinates module, the axes

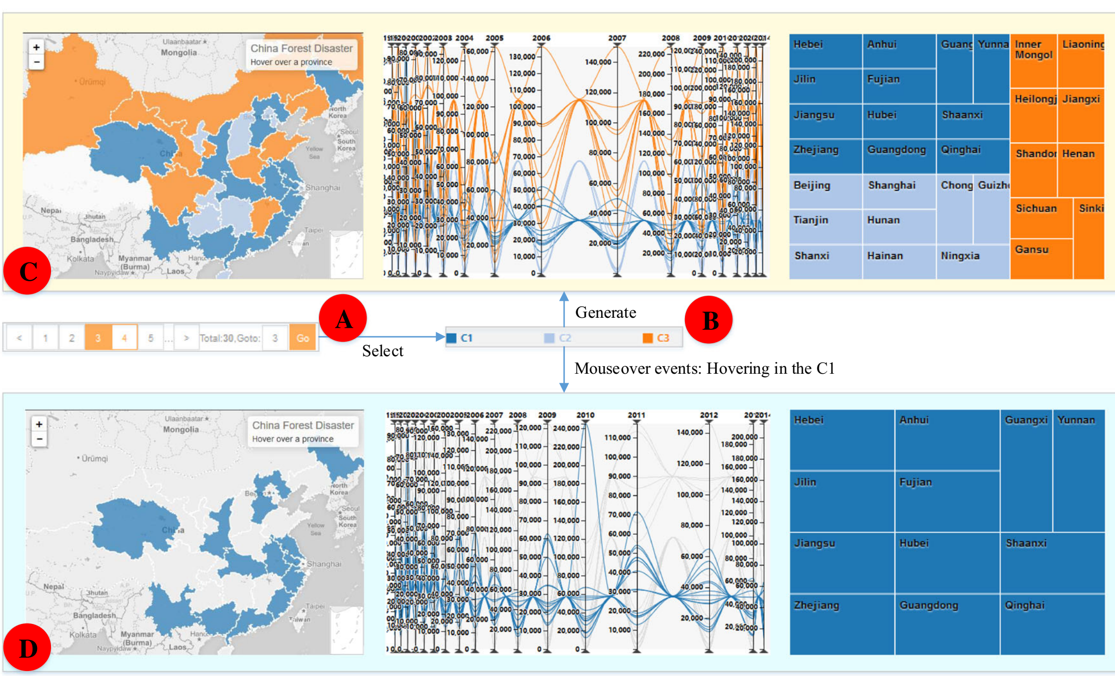

Fig. 13 Visual flow of cluster analysis. (**A**) Cluster number selector. (**B**) Categorical legends. (**C**) Visual display after clustering. (**D**) A visual display of category C1 when the mouse hovers over C1

represent all years while the lines represent the disease occurrences of each year. The parallel coordinates can display the situations every year on the screen where the clustering results can be arranged according to the cluster number beside the core point set between the axes as shown in Fig. 13C. The severity level of diseases of the cluster results increases along the axes from the bottom to the top. The system uses treemap to display which areas are contained in each category. In addition, the system adds a series of interactive operations. In the map module, a mouse hover over a specific area will highlight that area and induce the display of detailed information about the area on the upper right corner of the map. In the parallel coordinates module, axis exchange, brush and fisheye pattern are added. In the treemap module, the whole and local switching function is added. These interactive operations can facilitate the analysis of the cluster results.

Mouse hovers over the cluster legends in Fig. 13B arouse relevant views as shown in Fig. 13D. A mouse hover over C1 in the map module shifts the map to display only the cluster results of C1 for the users to more conveniently observe the regional distribution of the C1 category. Similarly in the parallel coordinates module, only the cluster results of C1 are also displayed for the users to observe the annual general trend of the category. In the treemap module, the display of treemap switches automatically from the whole to the corresponding local to show which areas are included in each cluster.

5 Case study

This section analyzes three cases to test the functions of the system with the data of forest diseases in China from 1998 to 2014. The effectiveness of the visualization method is verified in multiple analytical aspects such as the overall situations of forest diseases, forest diseases in a single region and the inter-regional similarities of forest diseases.

5.1 Case study 1: the analysis of overall disease situations

In the overall view module, the line chart in Fig. 2 shows that forest diseases increase generally in China in the selected 17 years. The diverging bar chart in Fig. 5 shows that there are both increases and decreases during the 17-year interval in which 2005 has the most severe annual increase of 33.8%. The multi-area chart shows further that among the three forest disease severity levels of which mild > moderate > severe, mild diseases increase while the moderate and severe diseases largely remain constant. Yet in 2005, diseases in all three levels increase as shown in Fig. 6. The map in Fig. 14 instinctively shows that the increase is mainly caused by significant regional increases in forest diseases in both Jilin and Jiangxi. The sortable table in Fig. 15 also shows that the increase in diseases in both regions is significant.

5.2 Case study 2: the analysis of diseases in a single region

This section takes the Shandong case as example to illustrate how the system analyzes forest diseases in a single region. In the map module, a click on Shandong pops up a window to show the overall situation of the region. The line chart shows that there are both increases and decreases in diseases between 1998 and 2005, the worst year after which a trend of gradual decrease appears. Graphs of a, b, c in Fig. 16 show diseases of different levels that all decease gradually after 2005. Detailed observations show that mild diseases have a linear trend similar to that of the overall situations. This indicates that the mild diseases account for a large percentage of the overall diseases. The sortable table in Fig. 17a shows that the region was the most affected one by forest diseases in the country in 2005. Graphs b, c, d in the sortable table 17 show that the region has high level of mild and moderate, but not severe disasters. Such analytical results inspire experts to inquire the reasons for such results.

5.3 Case study 3: the cluster analysis of regions

The inter-regional similarities and differences are analyzed with the cluster analysis module in the system. Since the data of 17 years are used with annual affected area as the featured dimension in the cluster algorithm, each region is represented with a 17-dimensional vector. The Euclidean distance of each vector is therefore calculated to obtain the final cluster results. This section illustrates the procedure by dividing the data into three categories. As shown in Fig. 13, a user first selects three as the cluster number (Fig. 11A) for

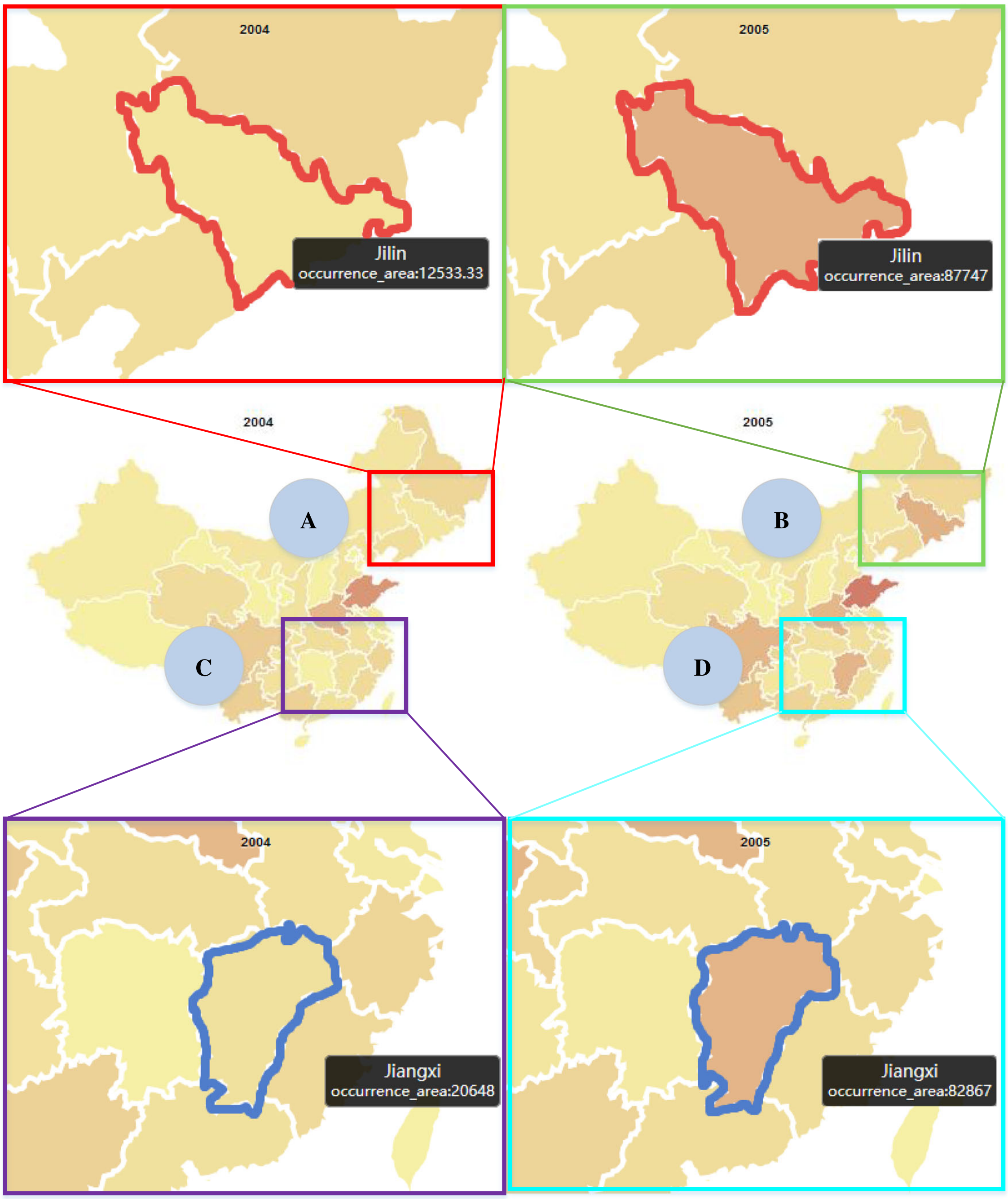

Fig. 14 **A** Severity of diseases in Jilin in 2004. **B** Severity of diseases in Jilin in 2005. **C** Severity of diseases in Jiangxi in 2004. **D** Severity of diseases in Jiangxi in 2005. 2005 is worse than 2004 for both regions

the system to automatically generate the corresponding identities of three categories, as shown in Fig. 13B. Thereafter, the system generates the corresponding cluster results to be displayed with the map, parallel coordinates and treemap modules, respectively, as shown in Fig. 13C. In the map module, geographical distributions of each category of forest diseases are displayed. The C3 category distributes in north China, the C1 category mostly concentrates in south China, while the C2 category concentrates in central China. In the parallel coordinates module, we can see that the degree of affectedness of disease categories ranks as C3 > C1 > C2. The treemap module displays the names of the regions in each category on the screen for the users to directly discern which areas have similar features. The cluster analysis module helps to find out

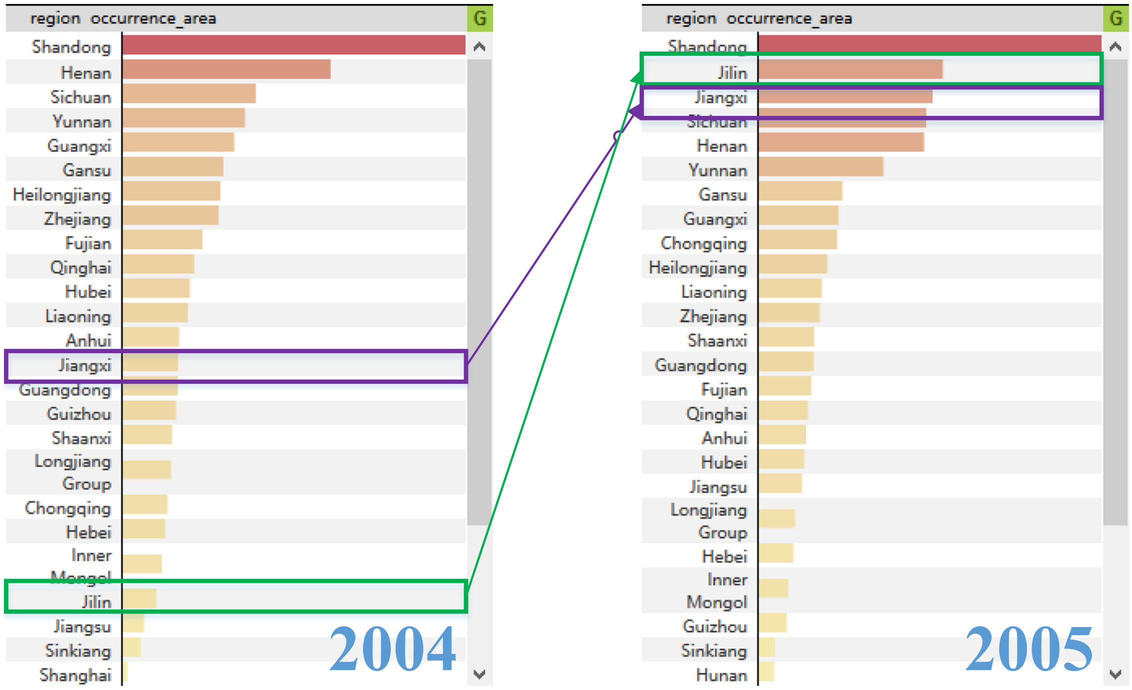

Fig. 15 Jiangxi Province rose from 14th in 2004 to 3rd in 2005. Jilin Province rose from 22nd in 2004 to 2nd in 2005

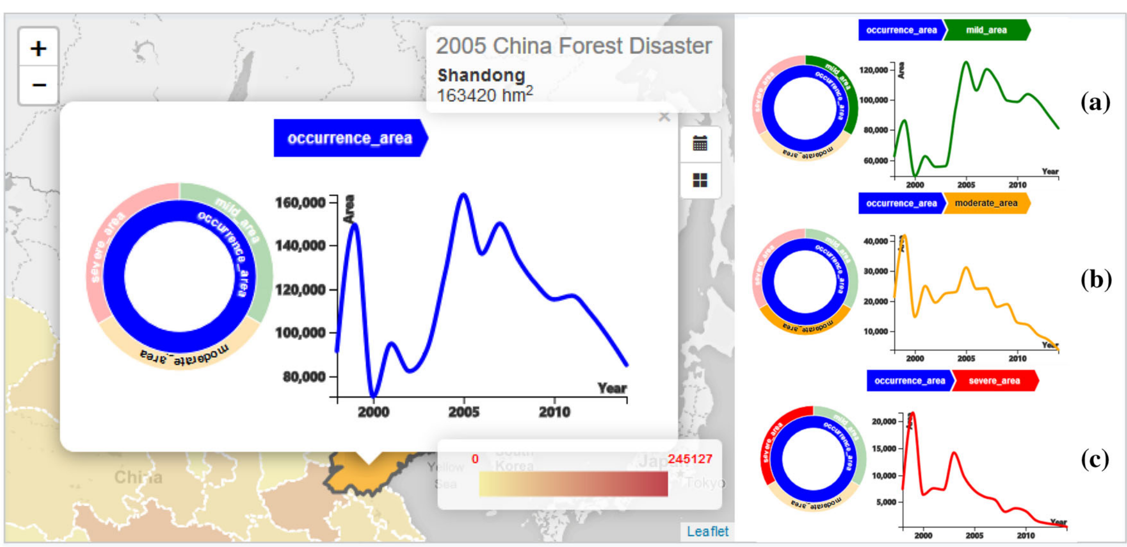

Fig. 16 Forest diseases in Shandong Province: the left graph shows the overall situation; the right graph shows the occurrence of various levels: **a** mild diseases, **b** moderate diseases, **c** severe diseases

regions affected by similar diseases so that successful disease prevention experiences in one region can be applied to peer regions.

6 Conclusion

We introduce an online visual analytic system that assists users to explore forest disease data interactively. The system is designed and realized on the basis of such technologies as time series visualization, map visualization, cluster analysis visualization and human–computer interaction. The effectiveness of our visual

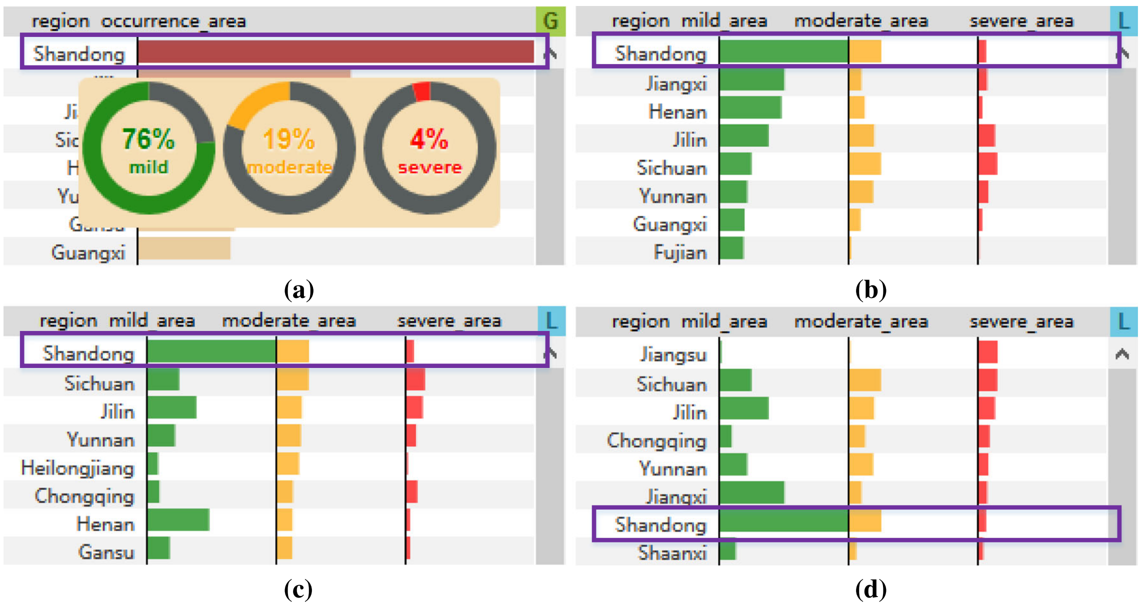

Fig. 17 Sortable table view of Shandong Province. **a** overall sort, **b** mild sort, **c** moderate sort, **d** severe sort

presentation approach has been confirmed by case studies and expert reviews. The system is user-friendly to researchers of forest diseases and is easy to be integrated with other systems. Nevertheless, there is room for further improvement in multiple aspects such as visualization optimization, interactive methods enrichment and system interface aesthetics. Our future research plan is to design more flexible visual presentation formats, incorporate more detailed disease information and make the system applicable to a broader range.

Acknowledgements The authors thank the reviewers for their comments. This work is supported by National Forestry Industry Research Special Funds for Public Welfare Projects (No. 201204503) and the Fundamental Research Funds for the Central Universities (No. 2015ZCQ-XX).

References

Aigner W, Miksch S, Schumann H, Tominski C (2011) In: Karat J, Vanderdonckt J (eds) Visualization of time-oriented data. Human-computer interaction series. Springer, London

Becker R, Ceveland A, William S (1987) Brushing scatterplots. Technometrics 29(2):127–142

Bureau S.F (1998–2014) China forestry statistics yearbook. China Forestry Press, Beijing

Chen Q, Xu Z, Zhang L, Lu P, Zhang Y (2018) Geostatistical analysis of the spatial distribution of arhopalus rusticus larvae and adults. Acta Ecol Sin 38(3):975–983. https://doi.org/10.5846/stxb201611182350

Chen W, Shen Z, Tao Y (2013) Data Visualization. Publishing House of Electronics Industry, Beijing

Chen Y, You Y, He Z, Zeng Y (2014) Analysis on temporal–spatial feature of forest fires in Chongqing. Chin Agric Sci Bull 13:51–58

Furnas GW (1986) Generalized fisheye views. Acm Sigchi Bull 17(4):16–23

Hartigan JA, Wong MA (1979) A *k* means clustering algorithm. Appl Stat 28(1):100–108

He K, Lu B (2018) Investigation and analysis of diseases and insect pests in fine variety base of *Pinus massoniana* Lamb. in Guizhou. Agric Technol 38(1):91–93. https://doi.org/10.11974/nyyjs.20180132038

Holten D (2006) Hierarchical edge bundles: visualization of adjacency relations in hierarchical data. IEEE Trans Vis Comput Gr 12(5):741–748. https://doi.org/10.1109/TVGC.2006.147

Holten D, Van Wijk JJ (2009) Force-directed edge bundling for graph visualization. Comput Gr Forum 28(3):983–990. https://doi.org/10.1111/j.1467-8659.2009.01450.x

Hu L, Zheng C, Yu S, Huang F, Zhang W (2013) Analysis of the species and distribution of taxus in Kaihua county in Zhejiang province. East China For Manag (02), 5–7+13

Hu Q, Hu T, Wang Y, Wang S, Cao K (2016) Survey on the occurrence and distribution of apple diseases in China. Plant Prot 42(01):175–179. https://doi.org/10.3969/j.issn.0529-1542.2016.01.032

Huang Y (2010) Analysis on the trend of major forest diseases and pests in Sanming city. Fujian Sci Technol Trop Crops 02:18–20

Inselberg A (1985) The plane with parallel coordinates. Vis Comput 1(2):69–91. https://doi.org/10.1007/bf01898350

Johnson B, Shneiderman B (1991) Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In: Proceedings of IEEE Conference on Visualization, 1991. Visualization '91, pp 284–291

Li S (2018) Types of fungal diseases in Guanmenshan region in Benxi Manchu autonomous county. Prot For Sci Technol (1), 54–56 https://doi.org/10.13601/j.issn.1005-5215.2018.01.017
Li Z, Wang X (2005) Eof analysis of spatial and temporal variation of forest fire in Zhejiang province. In: Chinese meteorological society annual meeting 2005, p 8
Liu S, Wu Y, Wei E, Liu M, Liu Y (2013) Storyflow: tracking the evolution of stories. IEEE Trans Vis Comput Gr 19(12):2436–45
Liu T (2016) The influence of forest disasters on the economic growth of forestry in China. China For Ind 2016(10):192
Liu X, Yang Z, Wang G, Li G (2016) Study on the main disease distribution of alfalfa in Heilongjiang province. Chin J Grassl 38(05):115–120. https://doi.org/10.16742/j.zgcdxb.2016-05-19
Mayinuer M, Zumina Sajida A, Guo Q (2017) Species and distribution of cotton root and stem diseases in eastern Xinjiang based on molecular detection. Xinjiang Agric Sci 54(11):2046–2053. https://doi.org/10.6048/j.issn.1001-4330.2017.11.011
Ocagne M (1885) Coordonnes parallles and axiales: mthode de transformation gomtrique et procd nouveau de calcul graphique dduits de la considration des coordonnes parallles. Nature 31:551. https://doi.org/10.1038/031551b0
Peng Q, Dong Z, Geng Y, Ma J (2003) Drawing forest pest distribution maps with computer. For Pest Dis 22(02):30–31. https://doi.org/10.3969/j.issn.1671-0886.2003.02.013
Plaisant C, Mushlin R, Snyder A, Li J, Dan H, Shneiderman B (2003) Lifelines: using visualization to enhance navigation and analysis of patient records. Proc Amia Symp 5(1):76–80
Rosling H (2006) The best stats you've ever seen http://www.ted.com
Sarkar M, Brown MH (1992) Graphical fisheye views of graphs. In: Sigchi conference on human factors in computing systems, pp 83–91
Shang X, Yao Y, Zhao W (2015) Geographic distribution of an invasive insect pest, *Obolodiplosis robiniae* (diptera: Cecidomyiidae) in China. For Pest Dis 2015(04):33–36. https://doi.org/10.3969/j.issn.1671-0886.2015.04.009
Tanahashi Y, Ma KL (2012) Design considerations for optimizing storyline visualizations. IEEE Trans Vis Comput Gr 18(12):2679–2688. https://doi.org/10.1109/TVCG.2012.212. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6327274
Torguson JS (2016) Choropleth map. Wiley, New York. https://doi.org/10.1002/9781118786352.wbieg0951
Wang TD, Plaisant C, Shneiderman B, Spring N, Roseman D, Marchand G, Mukherjee V, Smith M (2009) Temporal summaries: supporting temporal categorical searching, aggregation and comparison. IEEE Trans Vis Comput Gr 15(6):1049–56
Wen X, Li H, Kuang Y (1999) The situation of forest diseases and pests in Jiangxi province in 1998 and the analysis of the trend of pest situation in 1999. Jiangxi Plant Prot 02:4–6
Wilson JM (2003) Gantt charts: a centenary appreciation. Eur J Oper Res **149**(2):430–437. https://doi.org/10.1016/S0377-2217(02)00769-5. http://www.sciencedirect.com/science/article/pii/S0377221702007695
Xu X, Sun H, Li M (2018) Study on the distribution and prevention of forest pest in Hongshan district of Chifeng city based on gis. Inner Mong For Investigation Des **41**(01), 57–58+60 https://doi.org/10.13387/j.cnki.nmld.2018.01.021
Yang F, Zhang A, Meng F, Huo L, Li X, Di D, Miao H (2018) Distribution and genetic diversity of *Barley yellow striate mosaic virus* in northern China. Sci Agric Sin 51(2):279–289. https://doi.org/10.3864/j.issn.0578-1752.2018.02.008
Zhang PX, Zeng JP, Zeng C, Peng LH (2014) Status quo, losses and trend of forest diseases from 2000 to 2010 in China. Biol Disaster Sci 0(2):101–108. https://doi.org/10.3969/j.issn.2095-3704.2014.02.001