Bidirectional Spatio-Temporal Association Between the Observed Results of *Ulva* Prolifera Green Tides in the Yellow Sea and the Social Response in Sina Weibo

Zhongyuan Wang, Zhixiang Fang, Yu Zhang, and Zhanlong Song

***Abstract*—Massive green tides caused by *Ulva* prolifera have annually occurred in the Yellow Sea since 2007, which has attracted much attention from the government and society. There are associations between the green tides in the Yellow Sea and social response in the social media (i.e., Sina Weibo), which are bidirectional and could be captured by the bidirectional neural network. For instance, how to detect daily *U. prolifera* green tides by fusing remote sensing data with social media data, and how to use the observed *U. prolifera* green tides to infer the social response are two challenges of State Oceanic Administration, China. This article first illustrated that there are bidirectional associations between green tides and Sina Weibo data. Then, this article introduced a bidirectional spatio-temporal associative memory neural network (BSAMNN) model for modeling this bidirectional association from the spatio-temporal perspective. BSAMNN first extracted six characteristics from green tides and nine characteristics from the social responses in 2016–2019. Second, these characteristics were split by year, and the characteristics in 2016–2018 were, respectively, put into the bidirectional associative memory neural network (BAM), which is a two-layer artificial neural network. Based on the BAM results and the observed data, the residual network was constructed. Third, BSAMNN extracted the spatio-temporal rules from the characteristics in 2016–2018 as the constraints through the mining rule algorithm and modify the results via sea surface wind and ocean surface current. Last, BSAMNN put the characteristics in 2019 into BAM and used the residual network to modify the results, which was constrained by the spatio-temporal rules. The feasibility and reliability of our approach were demonstrated by using the *U. prolifera* green tides in 2019. The average accuracy, false alarm rate, and missing alarm rate of BSAMNN results were 0.69, 0.25, and 0.31, respectively, which was 0.11 higher, 0.10 lower, and 0.11 lower than that of the traditional BAM. The results indicated that our method is an effective alternative of linking the *U. prolifera* green tides and its public sentiments on social media.**

***Index Terms*—Green tide, remote sensing (RS), social media, spatial-temporal association, *Ulva* prolifera.**

Manuscript received January 15, 2021; revised March 2, 2021 and April 11, 2021; accepted May 27, 2021. Date of publication May 31, 2021; date of current version June 23, 2021. This work was supported in part by the National key Research and Development Plan under Grant 2017YFC1405302, and in part by the National Natural Science Foundation of China under Grant 41771473 and Grant 41231171. *(Corresponding author: Zhixiang Fang.)*

Zhongyuan Wang and Zhixiang Fang are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: whu_wzy@whu.edu.cn; zxfang@whu.edu.cn).

Yu Zhang is with the State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China (e-mail: zhangyu@ireis.ac.cn).

Zhanlong Song is with the Beijing Global Safety Technology Co., Ltd., Beijing 100010, China (e-mail: songzhanlong@gsafety.com).

Digital Object Identifier 10.1109/JSTARS.2021.3085090

I. Introduction

MASSIVE green tides, which are mainly caused by *Ulva* prolifera, have annually occurred in the Yellow Sea since 2007 [1]–[4]. These green tides have caused great harm to the coastal economy, maritime transportation, and tourism, which have attracted much attention from the government and society [5]–[7]. For instance, the green tides covered 13 000 km^2 during the 2008 Olympic Sailing Competition, and over 10 000 workers and 1400 boats were devoted to controlling the green tides [2], [8]. To effectively address the green tide issue, State Oceanic Administration (SOA) needs daily mapping on the coverage and distribution of green tides. At some time, social media platforms (i.e., Sina Weibo, which is China's largest social media platform) released so many opinions and feelings on the green tides, which could be used to estimate the development of green tides and social response. The government needs daily mapping results and the public sentiment data to develop an overall emergency response plan to minimize the losses and guide the development of public sentiment. For the task of daily mapping of green tides, remote sensing (RS) images (i.e., optical images and synthetic aperture radar (SAR) images) are the most popular data source. When RS images covered the Yellow Sea and were not contaminated by clouds, green tides could be effectively detected. When RS images are missing or seriously interfered by clouds, it is a big challenge for SOA to estimate the green tides on that day. For the public sentiment, they are usually later than the development of green tides, there is a challenge for the government to guide the public sentiment timely. Therefore, how to use daily mapping results and public sentiment data to build up a bidirectional association relationship of them to support the timely and daily emergency response is a big research question in the literature. If this bidirectional association relationship could be built up, the daily mapping results could help estimate the development of public sentiment, and the public sentiment data could help estimate the current development of green tides when RS images are missing. It could be very helpful for SOA to

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

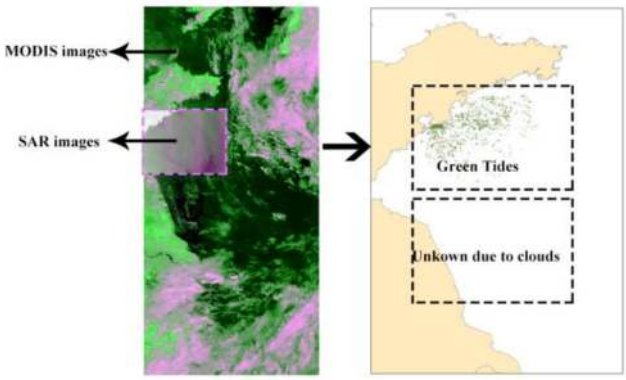


Fig. 1. MODIS image and SAR image of the Yellow Sea on July 11, 2019.

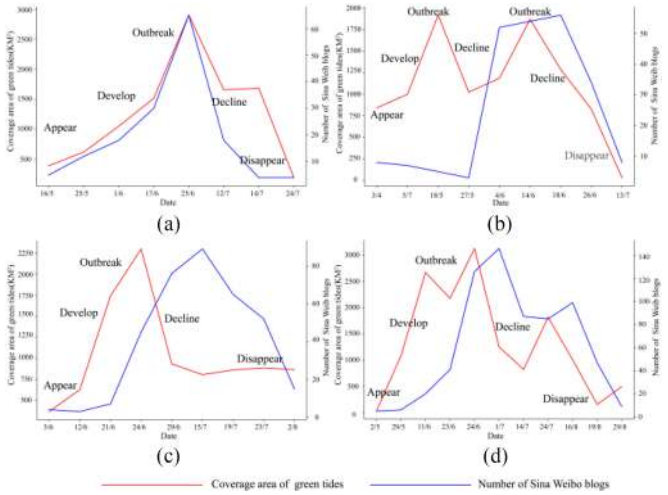


Fig. 2. CA and Sina Weibo blog number when green tide exists. (a) 2016. (b) 2017. (c) 2018. (d) 2019.

make a comprehensive emergency response plan for addressing at green tide issues.

Many studies have focused on the effective detection of green tides in the Yellow Sea [3], [5], [9]. Among these studies, optical images (i.e., MODIS images) and SAR images (i.e., Sentinel-1 images) are two popular data sources [2], [10]. MODIS images with a one-day repeat circle could meet the requirement of the daily report on *U. prolifera* green tides. However, as is shown in Fig. 1, there are always clouds in MODIS images, which will hinder the interpretation of the green tides. SAR images, which could eliminate the effect of clouds, are often used as then effective supplement to MODIS images. But the observation frequency and coverage of SAR satellite could not ensure the quality of the daily green tide report. As is shown in Fig. 1, MODIS images are covered by clouds. In Fig. 1, SAR images may not cover all areas with clouds. In current, many studies that use RS data to analyze the green tide tend to skip these limited situations [2], [3]. Between May and August in 2019, that is when *U. prolifera* green tides existed in the Yellow Sea, only 19 discrete days of MODIS images were available. During this period, there were only 26 days of discrete Sentinel-1 images with incomplete coverage of the Yellow Sea.

Apart from the effective detection of *U. prolifera* green tides, accurately assessing social response toward green tides is also an important task of SOA, which could help SOA identify potential affected areas and industries in advance. When green tides appear and develop in the Yellow Sea, many citizens will post their feelings and opinions on Sina Weibo, which can reflect the social response toward the green tides. As is shown in Fig. 2, this study found that the green tides and Sina Weibo blogs presented an approximate development trend in five stages (i.e., Appearance, Development, Outbreak, Decline, and Disappearance) [11]. From Fig. 2, it could also be found that the larger green tide scale tends to be associated with fewer lagging days. The phase lags of blog numbers behind the green tide coverage area (CA) in 2017, 2018, and 2019 are apparent, whereas that in 2016 is inapparent. The main reason to account for this phenomenon in 2016 may be the scale of green tides. Large-scale green tides may attract more citizens and departments to release information on Weibo. The peak value of green tides is 2906, 1919, 2301, and 3124 km^2, respectively, in 2016, 2017, 2018, and 2019, which means the green tide scale is larger in 2016 and 2019 than that in 2017 and 2018. The lagging days of blog number behind the peak of green tide area are 1, 31, 21, and 7, respectively, in 2016, 2017, 2018, and 2019, which means that the larger green tide scale may tend to be associated with fewer lagging days.

Furthermore, there may be bidirectional associations between green tides and Sina Weibo. Based on this assumption, the basic research question in this study is that does this assumption exists, if yes, how to build up a bidirectional spatio-temporal association model between the daily observation results of *U. prolifera* green tides in the Yellow Sea and the social response in social media data.

There are two contributions in this study. The first contribution is to illustrate the bidirectional associations between green tides and social response. This study extracted the CA, MBR, gravity center as the characteristics of green tides and extracted SDE, hot points and emotion as characteristics of Sina Weibo data. This study found that these characteristics showed similar trends in each year with the development of green tides. The second contribution is the introduction of the bidirectional spatio-temporal associative memory neural network (BSAMNN) model. This study first constructed a residual network between observed data and results from the bidirectional associative memory neural network (BAM). Then, the spatio-temporal constraints in *U. prolifera* green tides and Sina Weibo data were to be constructed. Finally, BSAMNN was constructed by improving BAM with the approaches of residual network and spatio-temporal constraints. These associations represented by BSAMNN could help SOA map the *U. prolifera* green tide when RS data quality is poor, and estimate the situation of public sentiments in social media for the current stage of *U. prolifera* green tides.

The article is organized as follows. Section II discusses the previous research related to *U. prolifera* green tide detection, as well as the estimation of the social influence of disasters based on Sina Weibo. Section III introduces the summarized characteristics of *U. prolifera* green tides and Sina Weibo data,

and illustrates the bidirectional associations between green tides and social response. Then, Section III introduces the BSAMNN. Section IV introduces the experiment results and discussion. Finally, Section V provides the conclusion and discusses future research.

II. Related Research Works

This section introduces the research works related to the study of *U. prolifera* green tide detection and the estimation of the social influence of disasters based on Sina Weibo.

A. U. Prolifera Green Tide Detection

The *U. prolifera* green tide detection is an essential precursor of emergency response to the annual green tides in the Yellow Sea. Its purpose is to monitor the distribution, CA, and morphological characteristics of the daily green tide. Research works have applied various methods to daily detection of *U. prolifera* green tides, such as SAR images [5], [12], optical images [13], [14], ocean surveillance ships, and ocean surveillance planes [15], [16]. Among these methods, MODIS images and SAR images are the most popular data source, which were used to quantify the coverage and affected areas of *U. prolifera* green tides [2], [10].

The most commonly used detection algorithm on RS images is the vegetation indexes, including the normalized difference vegetation index (NDVI) [17] and the floating algal index (FAI) [9]. Recently, Cao *et al.* [18] used NDVI calculated from MODIS to analyze the patterns and characteristics of *U. prolifera*. Hu proposed the FAI to detect large-surface-floating algae from MODIS 250 and 500m bands in both fresh and marine environments [9], [19] and FAI has been used in the *U. prolifera* green tide detection [3].

In the above research works, the detection of *U. prolifera* green tide has become accurate. However, the above studies have a strong dependence on RS images and marine environment data, which may be affected by clouds, repeat circles, and coverage of satellites. For example, MODIS images (one-day repeat cycle and 250-m resolution), Sentinel-2 images (ten-day repeat cycle and 10-m resolution) are two useful data source to monitor the *U. prolifera* green tides. MODIS has a notably higher observation frequency than Sentinel-2, which means the daily coverage of Seintiel-2 is smaller than that of MODIS images. But the lower spatial resolution of MODIS images compared with the Sentinel-2 images means that MODIS could only map the *U. prolifera* in a rough way [3]. In other words, it is difficult to effectively estimate and detect the *U. prolifera* green tide with the above methods in the case of missing and poor RS and marine environment data. Current data fusion approaches for detecting *U. prolifera* green tide seldom include social media data in the literature.

B. Estimation of Social Influence of Disasters Based on Sina Weibo

In recent years, the widespread use of the Internet and social media by the public, such as Sina Weibo, has made the detection and estimation of natural disasters quickly and at low cost, such as earthquakes, floods, and typhoons [20], and also support the air pollution monitoring, human mobility pattern detection [7], [11], [21]. Among the massive social media information, there are posts with geo-information, which could help researchers map the related disasters [22]. Meanwhile, the geo-distribution of posts has an association with the disasters. Many researchers have focused on the spatio-temporal patterns of social media posts, which could help efficiently and effectively acquiring disaster situational awareness information [22].

The work of Poblete *et al.* [23] shows that the massive social media data (i.e., Twitter) is an inexpensive and viable alternative to describe the disasters timely and effectively. Their methods are based on a one-off semisupervised initialization and could be scaled to track multiple signals in a robust manner. The work of Earle *et al.* [24] shows that it is possible to use social medial data for qualitative assessment and the rapid detection of earthquakes. Their method builds the frequency and time series of social media data to identify earthquakes as much as possible, which relies on Twitter data entirely. Besides the algorithms above, statistic classifiers are commonly used in indicating the occurrence of disasters. For example, the work of Imran *et al.* [25] trained naïve Bayesian classifiers, symbols (i.e., reference), tweet length, part of speech tags, and other characteristics. In general, there are mature methods for detecting natural disasters such as earthquakes via social media. However, there are no green-tide detection methods via social media data and also no studies of social media having been conducted to predict the spatio-temporal dynamics of *U. prolifera* green tide.

III. Methodology and Material

Fig. 3 briefly illustrates the research content of this part. After the data preprocessing of the green tides and social medial data, this study first extracted the CA, gravity center, and minimum area bounding rectangle (MBR) as the characteristics of the green tides, the standard deviation ellipse (SDE), emotion and the spatio-temporal points as the characteristics of social media data. Through these characteristics above, this study illustrates that there are bidirectional associations between green tides and Sina Weibo data. Then, the bidirectional association between the characteristics of the green tides and the social media data were constructed based on the BSAMNN model. Finally, the accuracy, the false alarm rate (FAR), and the missing alarm rate (MAR) were used to estimate the experiment results.

A. Data Preprocessing

1) Green Tides' Data Preprocessing: The Yellow Sea is a shallow marginal sea located between China and Korea. It connects to the Bohai on the north and is separated from the western Pacific Ocean by the East China Sea to the South. The annual variation in water temperature is 13 °C–24 °C and the seawater salinity is about 33‰, which is suitable for the *U. prolifera* development. Since 2007, the green tides have erupted regularly for the past ten years. This study used the *U. prolifera* green tides in the Yellow Sea from 2016 to 2019 as the experimental data. Level-1B 250m

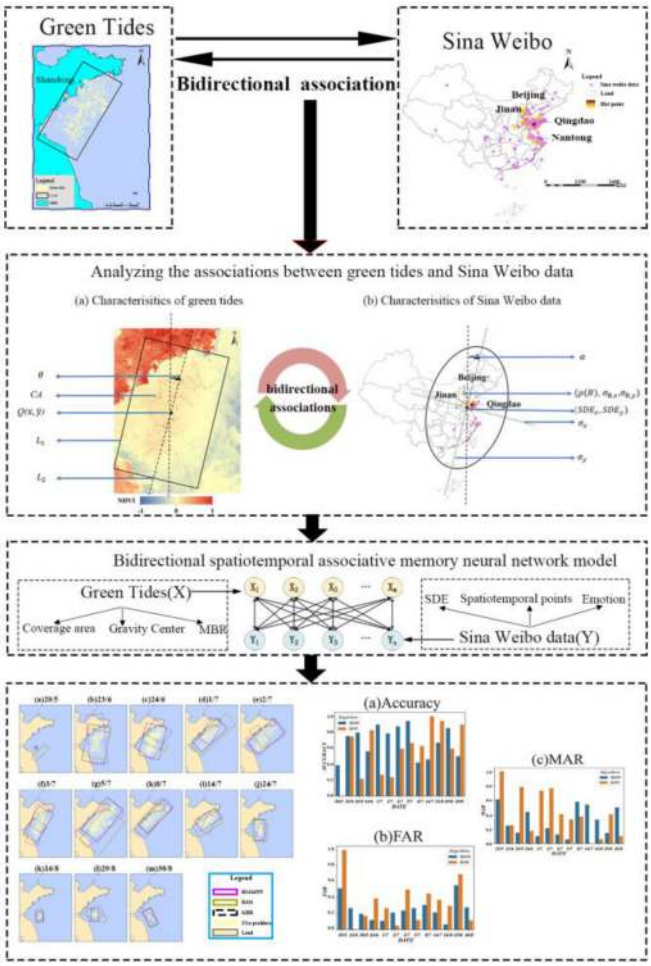


Fig. 3. Framework of this research.

TABLE I
MOD02QKM SELECTED IN THIS STUDY

Year	Date
2016	May 16, May 25, June 1, June 17, June 25, July 12, July 14, July 24
2017	April 30, May 7, May 18, May 27, June 4, June 14, June 18, June 26, July 13
2018	June 3, June 13, June 21, June 24, June 29, July 15, July 19, July 23, August 2
2019	May 20, May 22, May 23, May 29, June 7, June 11, June 15, June 23, June 24, July 1, July 2, July 3, July 5, July 14, July 24, August 16, August 19, August 29, August 30

class data (MOD02QKM) of MODIS developed by NASA (https://ladsweb.modaps.eosdis.nasa.gov/search/) were used to extract the *U. prolifera* green tides. MOD02QKM covers red and near-infrared (NIR) bands that can be used to discriminate clouds, plants, and land, which means that MOD02QKM helps obtain information about *U. prolifera*. This study selected MOD02QKM with no clouds from April to August 2016–2019 in the Yellow Sea as the data source, which is presented in Table I. It is worth noting that SAR images (i.e., Sentinel-1 images) are an important data source to detect the *U. prolifera* green tides, especially when there are clouds in MODIS images. But the frequency and coverage of SAR images could not meet the

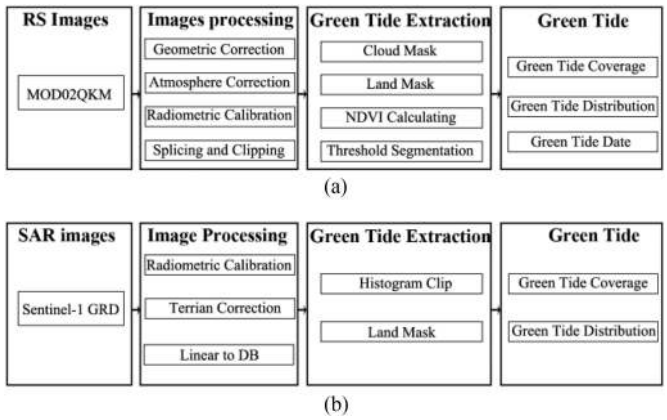


Fig. 4. Preprocess of green tides. (a) Extracting green tides using MODIS images. (b) Extracting green tides using SAR images.

TABLE II
SENTINEL-1 GRD PRODUCTS IN 2019 SELECTED IN THIS STUDY

Month	Date
May	23、 24 、29、 31
June	5、 10、 12、 17、 22 、24、 29
July	4 、 6 、11 、16、 18、 23 、28 、30
August	4 、 9 、11、 16 、21、 23、 28

requirements of daily *U. prolifera* mapping [3]. As is presented in Table I, the MODIS images that could be used in the green tide detection are not continuous in time. The purpose of this study is to estimate the *U. prolifera* green tides in the absence of RS images based on the social media data.

Fig. 4 illustrates the preprocess of the *U. prolifera* green tide data. As is shown in Fig. 4(a), to obtain the *U. prolifera* green tides in the Yellow Sea, MODIS images (i.e., MOD02QKM) that are most popular in *U. prolifera* research were first downloaded. Then, geometric correction, atmosphere correction, radiometric calibration, image splicing, and image clipping were conducted in the Environment for Visualizing Images. After masking the clouds and lands in images, the *U. prolifera* green tides in the images were obtained by calculating the NDVI [17]. Considering that floating algae have a higher reflectance in the NIR bands than the seawater and low reflectance in the red bands, the NDVI formula evaluates this difference as follows:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}} \tag{1}$$

where NIR and RED represent the near-infrared and red bands, respectively. The image pixels where NDVI is positive indicated that there is *U. prolifera* coverage, which is the basis of the threshold segmentation. The coordinate system set in this study was WGS_1984_UTM_ZONE_51N. More details about how to extract the green tides via MODIS images could be found in the work of Cao *et al.* [2].

When there are clouds in MODIS images, SOA will prefer to using SAR images (i.e., Sentinel-1 images) to detect the *U. prolifera* green tides. Table II presents the date of Sentinel-1 GRD product, which covered the Yellow Sea from May to August in 2019. Fig. 4(b) shows the process of extracting the green tides using SAR images. After acquiring the Sentinel-1 GRD products

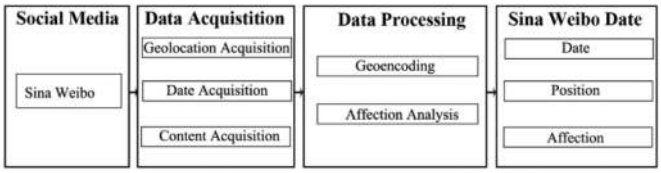


Fig. 5. Preprocess of Sina Weibo data.

on the date shown in Table II, this study conducted radiometric calibration, terrain correction, and linearization to dB in SNAP 8.0. To compose the RGB images, the cross-polarized channel (VH) was used as red and blue channel and the copolarized channel (VV) was used as the green channel. After land masking and histogram clipping, the *U. prolifera* green tides will be obtained in the Yellow Sea, which could be used as an effective complement to MODIS images. More details about how to extract the green tides via SAR images could be found in the work of Cui *et al.* [5].

2) Sina Weibo Data Preprocessing: Fig. 5 illustrates the preprocess of the Sina Weibo data. The Sina Weibo data employed in this study was collected via the interface of the Sina-Weibo API and their text messages were all related to the green tides in the Yellow Sea. This study took green tides, *U. prolifera*, green tide disaster, green tide disposal, green tide salvage, marine grassland, and *U. prolifera* drift as keywords of the search.

This study first extracted the text message, a geo-location, and the release date in each Sina Weibo messages. Then, this study set the coordinate system of the geo-location WGS_1984_UTM_ZONE_51N to make it consistent with green tide data. This study used the SnowNLP platform (https://github.com/isnowfy/snownlp) to explore the emotion of each message. The value of emotion is between 0 and 1, with 0 indicating a strong negative mood and 1 indicating a strongly positive mood. This study extracted 1084 effective messages from Sina Weibo. It is worth noting that there were 63% messages with positive mood. Although the *U. prolifera* green tides do harm to the maritime traffic and maritime aquaculture, many people view the green tides as landscape and post their positive mood on Sina Weibo. In addition, some users will post positive feelings when green tides declines.

3) Sea Surface Wind (SSW) Data and Ocean Surface Current Preprocessing: This study assumed that *U. prolifera* green tides in the Yellow Sea have a bidirectional association with Sina Weibo data. Based on this assumption, this study could use Sina Weibo data to predict the green tides in the Yellow Sea. It is worth noting that the green tide distribution is affected by many physical factors such as SSW and ocean surface current [26]. For example, green tides move from Jiangsu to Shandong annually due to SSW. To ensure the algorithms and models proposed by this study to be more realistic, this study used WindSat data and OSCAR third-degree resolution ocean surface current data from May to August in 2016–2019 to constrain the results of the BSAMNN. OSCAR third-degree resolution ocean surface current data could be download from https://doi.org/10.5067/OSCAR-03D01 [27]. The zonal velocity (U) and meridional velocity (V) were used to merge the ocean surface currents in the Yellow Sea. The ocean surface current data used in this study are shown in Fig. 6. WindSat data are produced by Remote Sensing Systems (http://www.remss.com/missions/windsat/) and sponsored by the NASA Earth Science MEaSUREs DISCOVER Project and the NASA Earth Science Physical Oceanography Program. RSS WindSat data are available at www.remss.com. This study extracted the all-weather 10-m surface wind speed and surface wind direction as SSW from WindSat data. Fig. 7 shows the SSW data extracted from WindSat. In the Yellow Sea, the average SSW speed in May, June, July, and August are 4.46, 4.45, 5.28, and 6.38 m/s, respectively. The average SSW direction (relative to north) is 191.78°, 224.24°, 230.06°, and 212.89°, respectively.

B. Spatial Indicator and Pattern of U.Prolifera Green Tides and Sina Weibo Data

1) Green Tide CA: The definition of green tide coverage in this study is the area with explicit *U. prolifera* coverage in MOD02QKM images. The green tide CA is one of the most important indexes of SOA when the green tides appear in the Yellow Sea, which can be calculated as follows:

$$\mathrm{CA} = N_{\mathrm{NDVI}>0} \cdot \mathrm{cellsize}^2 \tag{2}$$

where $N_{\mathrm{NDVI}>0}$ represents the pixel number of MODIS images whose NDVI is positive and cellsize means the spatial resolution of the satellite. In this study, cellsize is 250 m. It is worth noting that many pixels in the MODIS images are not completely occupied by the green tides, for any pixel that contains even a little *U. prolifera*, the area of the whole pixel is included in the calculation [3], [5].

In the existing literature, the development of *U. prolifera* green tides in the Yellow Sea is summarized as five major stages: appearance, development, outbreak, decline, and disappearance [2], [11]. The appearance stage is the first detection of *U. prolifera* green tides with CA less than 10 km^2. The outbreak stage is the period when CA is over 1500 km^2. The stage between appearance and outbreak is called the development stage. After the outbreak stage, the CA of green tides is gradually declined. The disappearance stage is the period when it is difficult to detect the *U. prolifera* in RS images.

2) Gravity Center of Green Tides: The gravity center of green tides reflects the central position of *U. prolifera*. Many studies have concluded that the green tides in the yellow Sea drift from the coastal waters of the Jiangsu and drifted to the southern coast of the Shandong due to the SSW and ocean currents [2], [9], [28]–[30]. Suppose that the green tide on a certain day distributed at *n* points *P*, and the gravity center of green tides can be described as $Q(\bar{x}, \bar{y})$

$$\begin{cases} \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \\ \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \end{cases}. \tag{3}$$

3) MBR of Green Tides: The distribution of *U. prolifera* green tide is one of the most important significant characteristics. In the previous research works, the outer contour of *U. prolifera* covering pixels on satellite images is chosen as the green tide distribution in many research works [2], [29], [31]. As is shown

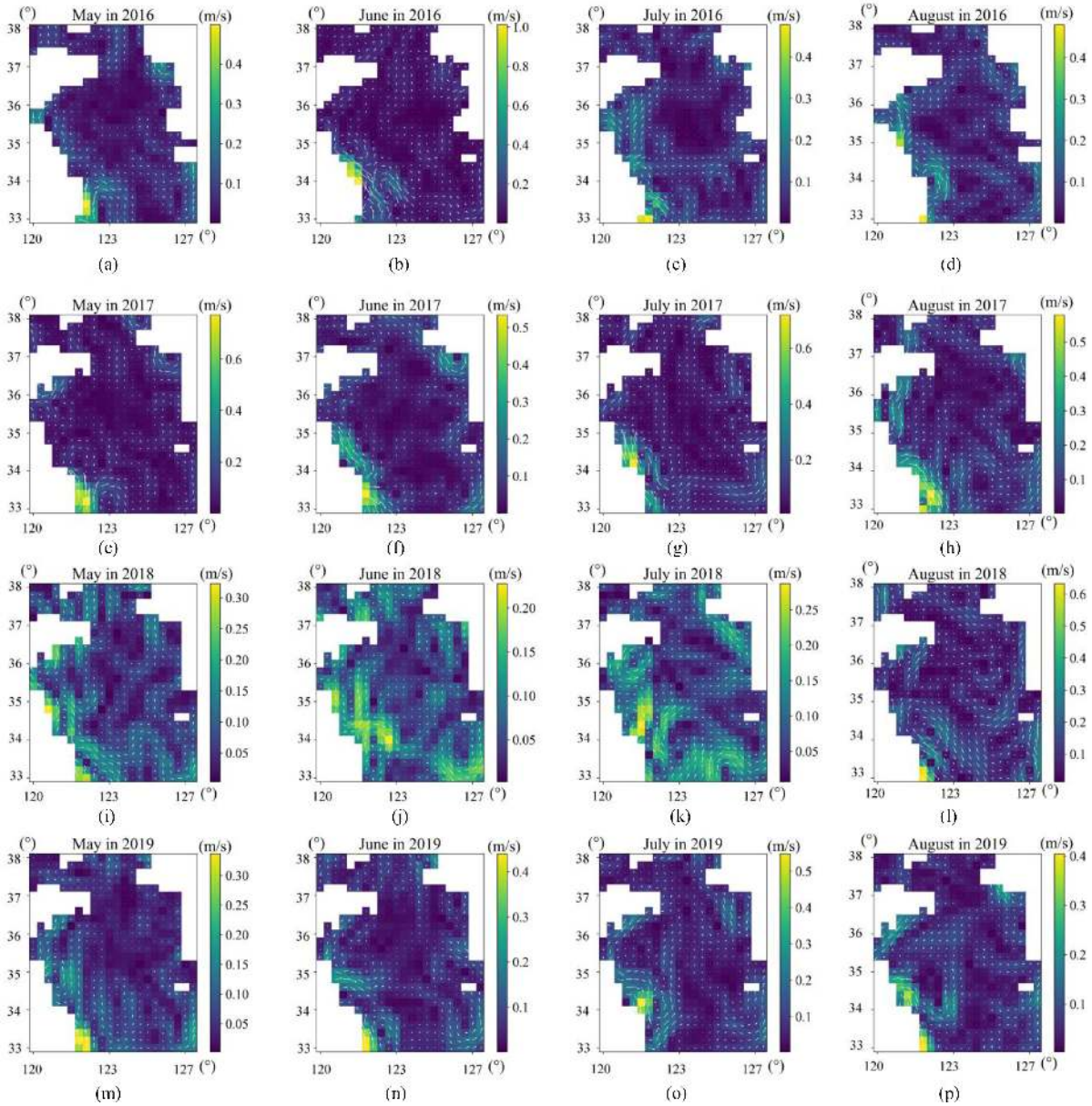


Fig. 6. Ocean surface current data used in this study.

in Fig. 8(b) and (c), the contour of *U. prolifera* coverage pixels could clearly express the distribution pattern and size of green tides. But the contour is not unique because it is affected by the distance threshold [31], which means the green tides at a certain day do not correspond to only one outer contour. The MBR of the *U. prolifera* coverage, as is shown in Fig. 8(a) is the only correspondence to the certain *U. prolifera* coverage, which is the first reason accounting for MBR rather than the counter is used as the distribution in this study.

Second, the irregular polygon distribution of green tides could be extracted based on the pixels with green tide on RS images. Compared with RS images, it is obvious that the ability of social media data to characterize pixels with green tide is inadequate. Without the pixels with green tide, the irregular polygon could not be formed. Compared with irregular polygons, MBR could be uniquely described by the width, length, and angle, which is another reason accounting for MBR rather than the counter is used as the distribution in this study. Third, the ability to describe green tides' position of distribution is not as accurate as that of coverage. To describe the impact range of green tides, MBR is suitable.

Suppose the green tide on a certain day distributed at n points P and the $(x_i,\ y_i)$ is the coordinates of the ith point. Let $X = \{x_i | i = 1, 2, \ldots, n\}$, $Y = \{y_i | i = 1, 2, \ldots, n\}$. To obtain the MBR of P, this study used the calculation steps as follows.

Step 1: Calculate the Range(R) of X and Y as two sides of the boundary rectangle (BR), respectively.

Step 2: Rotate P clockwise by θ around $Q(\bar{x}, \bar{y})$ to obtain $P'(X', Y')$.

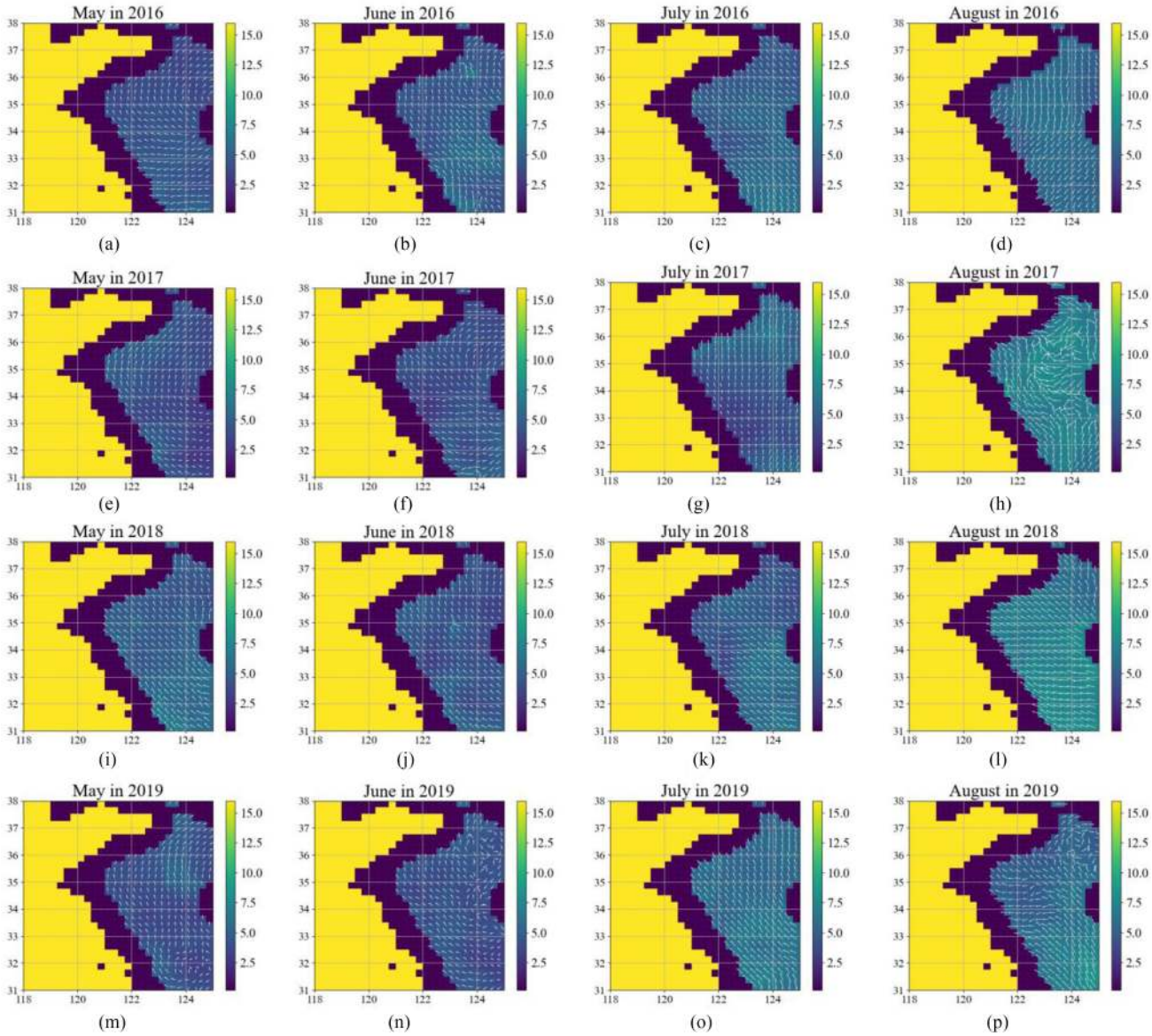


Fig. 7. SSW data used in this study.

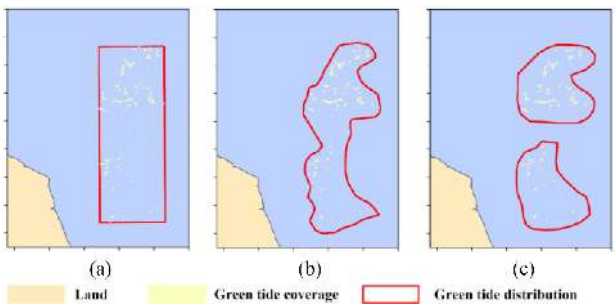


Fig. 8. Expression of green tide distribution.

Step 3: Calculate the Range(R) of X' and Y' as two sides of BR, respectively, and rotate BR counterclockwise by θ.

Step 4: Change θ and repeat Steps 2–3 and find the BR with the minimum area as MBR. Equation (4) shows the optimization function of MBR. In this study, θ is between 0° and 180°

$$\min F \quad (X, Y, \theta) = R\left[(X-\bar{x})\cos\theta + (Y-\bar{y})\sin\theta\right] R\left[(Y-\bar{y})\cos\theta - (X-\bar{x})\sin\theta\right]. \quad (4)$$

To describe the MBR, this study used θ, L_1, and L_2 as the spatial characteristic of MBR

$$\begin{cases} L_1 = R\left[(X-\bar{x})\cos\theta + (Y-\bar{y})\sin\theta\right] \\ L_2 = R\left[(Y-\bar{y})\cos\theta - (X-\bar{x})\sin\theta\right] \end{cases}. \quad (5)$$

4) Morphological Characteristic Index of Sina Weibo Data: The morphological characteristic of Sina Weibo Data is the overall distribution of the published location. This study used the SDE as the morphological characteristic of Sina Weibo Data. SDE can reflect the dominant distribution of the spatial elements (i.e., Sina Weibo data) [18], [32]. The key formula of SDE of Sina Weibo data on a certain day is given in

(6), as shown at the bottom of this page

$$\sigma_x = \sqrt{\frac{2\left(\sum_{i=1}^{n}(\tilde{x}_i \cos\alpha - \tilde{y}_i \sin\alpha)\right)}{n}} \tag{7}$$

$$\sigma_y = \sqrt{\frac{2\left(\sum_{i=1}^{n}(\tilde{x}_i \cos\alpha + \tilde{y}_i \sin\alpha)\right)}{n}} \tag{8}$$

$$\mathrm{SDE}_x = \frac{\sum_{i=1}^{n} x_i}{n} \tag{9}$$

$$\mathrm{SDE}_y = \frac{\sum_{i=1}^{n} y_i}{n} \tag{10}$$

where α is the angle of SDE rotation, σ_x and σ_y are the standard deviations for the *x*-axis and *y*-axis, (x_i, y_i) represents the spatial location of the objects, $(\tilde{x}_i, \tilde{y}_i)$ is the coordinate deviation from the location to the mean coordinate, and *n* is the total number of Sina Weibo data.

5) Spatio-Temporal Point Patterns of Sina Weibo Data: In the existence of *U. prolifera* green tides, many people share their emotions and opinions with others through online communities, which could have an impact on other people [33] and presents some forms of spatial aggregation. To describe the spatial aggregation of Sina Weibo data, this study used Poisson point process (PPP) [34], [35], which is the most widely spread point process model [36], [37]. Details of this PPP can be found in the literature [38], [39].

Let $s_1(x_1, y_1)$, $s_2(x_2, y_2), \ldots, s_n(x_n, y_n)$ denote event locations, the location of Sina Weibo data, let *D* denote the study area, and let *T* denote the study period. The event locations are viewed as the realization of PPP, whose intensity function $\boldsymbol{\lambda}$ is defined as follows:

$$\boldsymbol{\lambda}\ (s_i, t_i) = \lim_{|ds||dt|} \left\{ \frac{E\left[N\left(ds, dt\right)\right]}{|ds|\,|dt|} \right\} \tag{11}$$

where *N* is the number of events, and t_i is the *i*th part of *T*.

PPP model used in this study is defined by the following two conditions.

1) Any bounded borel set $B \subset D$ has $\rho(B) \in (0, \infty)$ and $N(B) \sim \mathrm{Poisson}(\rho(B))$

$$\rho\ \ (B) = \int_0^{t_i} \boldsymbol{\lambda}\left(s_i, t_i\right) dt. \tag{12}$$

2) The event locations $s_1(x_1, y_1)$, $s_2(x_2, y_2), \ldots, s_n(x_n, y_n)$ are related to the *B* and independent identically distributed in a bivariate normal distribution

$$h\left(x_n, y_n\right) = \frac{1}{2\pi\sigma_{B,x}\sigma_{B,y}}\ e^{-\frac{1}{2}\left(\frac{\left(x_n-\mu_{B,x}\right)^2}{{\sigma_{B,x}}^2}+\frac{\left(y_n-\mu_{B,y}\right)^2}{{\sigma_{B,y}}^2}\right)} \tag{13}$$

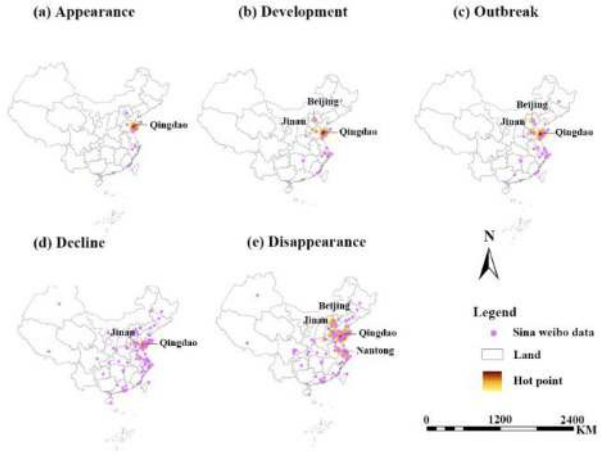


Fig. 9. Hot points of Sina Weibo data in 2016–2019. (a) Appearance. (b) Development. (c) Outbreak. (d) Decline. (e) Disappearance.

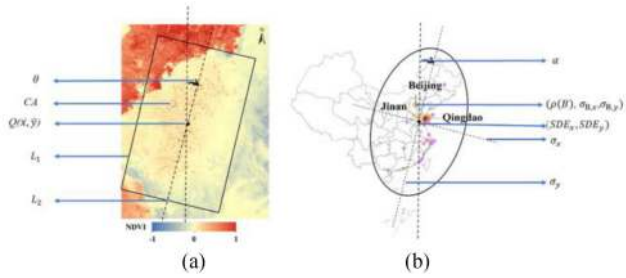


Fig. 10. Characteristics of (a) Green tides and (b) Sina Weibo data on August 19, 2019.

where $(\mu_B,\ \mu_B)$ represents the mean difference between the center point coordinate of *B* with s_n and $(\sigma_{B,x},\ \sigma_{B,y})$ represents the standard deviation of the distance between s_n and the center of *B*.

To describe and estimate the spatial distribution of the Sina Weibo data, this study used $\rho(B)$, σ_{B,x_n}, σ_{B,y_n}, μ_{B,x_n}, and μ_{B,y_n} as the characteristic of spatio-temporal point patterns of Sina Weibo Data. It is worth noting that this study used the hot point analysis tools in ArcGIS [32] and finds four cities (i.e., Beijing, Jinan, Qingdao, and Nantong) as *B*. As is illustrated in Fig. 9, Sina Weibo data are mainly gathered near Qingdao during the appearance of green tides. In the development and outbreak period of green tides, Jinan, Qingdao, and Beijing are three distribution centers of Sina Weibo data. When green tides decline, the centers are Jinan and Qingdao. When green tides disappear, Beijing, Jinan, Qingdao, and Nantong are four hot centers of Sina Weibo data.

Fig. 10 gives an example of the characteristics of green tides and Sina Weibo data. Through the NDVI in the Yellow Sea on August 19, 2019, CA, θ, L_1, L_2, and $Q(\bar{x}, \bar{y})$ of green tides can be calculated [see Fig. 10(a)]. On August 19, 2019, the green

$$\tan\alpha = \frac{\left(\sum_{i=1}^{n}\tilde{x}_i^2 - \sum_{i=1}^{n}\tilde{y}_i^2\right) + \sqrt{\left(\sum_{i=1}^{n}\tilde{x}_i^2 - \sum_{i=1}^{n}\tilde{y}_i^2\right)^2 + 4\left(\sum_{i=1}^{n}\tilde{x}_i\tilde{y}_i\right)^2}}{2\sum_{i=1}^{n}\tilde{x}_i\tilde{y}_i} \tag{6}$$

TABLE III
PEARSON CORRELATION ANALYSIS BETWEEN EACH TWO YEARS ($P<0.05$)

characteristic	Pearson correlation coefficients calculated between each two years					
	(2016,2017)	(2016,2018)	(2016,2019)	(2017,2018)	(2017,2019)	(2018,2019)
CA	0.4669	0.8430	0.7149	0.3153	0.5060	0.6704
L_1	0.7199	0.8424	0.8548	0.4401	0.5584	0.8587
L_2	0.6588	0.7989	0.8347	0.3432	0.4536	0.8577
θ	0.8414	0.6352	0.5952	0.5331	0.4816	0.5610
$\bar{x}$	0.6809	0.7866	0.7729	0.3977	0.4518	0.7437
$\bar{y}$	0.7188	0.8340	0.7693	0.5317	0.5205	0.7206
SDE_x	0.5514	0.2160	-0.0064	0.0618	0.0353	0.7708
SDE_y	0.6247	0.5206	0.7349	0.1610	0.4679	0.7312
σ_x	0.4706	0.4531	0.5470	0.2036	0.4923	0.6975
σ_y	0.3537	0.4665	0.3303	-0.0393	0.5190	0.3250
α	0.3623	0.3521	0.5217	0.2845	0.4055	0.8174
$\sigma_{\mathrm{B,x}}(B = Beijing)$	0.5219	0.3922	0.4420	0.0741	0.4723	0.3967
$\sigma_{\mathrm{B},y}(B = Beijing)$	0.4054	0.5514	0.5484	0.0148	0.6808	0.3946
$\sigma_{\mathrm{B,x}}(B = Jinan)$	0.5219	0.3922	0.4420	0.0741	0.4723	0.3967
$\sigma_{\mathrm{B},y}(B = Jinan)$	0.4054	0.5514	0.5484	0.0148	0.6808	0.3946
$\sigma_{\mathrm{B,x}}(B = Qingdao)$	0.5219	0.3922	0.4420	0.0741	0.4723	0.3967
$\sigma_{\mathrm{B},y}(B = Qingdao)$	0.4054	0.5514	0.5484	0.0148	0.6808	0.3946
$\sigma_{\mathrm{B,x}}(B = Nantong)$	0.5219	0.3922	0.4420	0.0741	0.4723	0.3967
$\sigma_{\mathrm{B},y}(B = Nantong)$	0.4054	0.5514	0.5484	0.0148	0.6808	0.3946
$\rho(B)(B = Beijing)$	0.3181	0.4405	0.4160	0.0492	0.0323	0.7748
$\rho(B)\ (B = Jinan)$	0.3177	0.4318	0.4108	0.0404	0.0265	0.7743
$\rho(B)\ (B = Qingdao)$	0.3198	0.4447	0.4175	0.0508	0.0257	0.7786
$\rho(B)\ (B = Nantong)$	0.3424	0.4806	0.3940	0.0308	0.1328	0.7073

tides were in the outbreak stage, on which the hot points of Sina Weibo data were Beijing, Qingdao, and Jinan [see Fig. 10(b)]. Through the Sina Weibo data, $\alpha, \sigma_x, \sigma_y, \rho(B), \sigma_{\mathrm{B},x}, \sigma_{\mathrm{B},y}$, and $(\mathrm{SDE}_x, \mathrm{SDE}_y)$ can be calculated. After extracting the characteristics of the green tides and Sina Weibo data, the bidirectional associations will be illustrated and established.

C. Illustration of the Bidirectional Associations Between the Green Tides and Social Response

This section first extracted the characteristics from *U. prolifera* green tides and corresponding Sina Weibo data in 2016–2019. Then, trends of each characteristic in each year were to be analyzed. This section used Pearson correlation analysis, which is a common algorithm to evaluate the similarity of two time series [40], to denote the similarity of the trends of green tides and social responses in each two years.

Pearson correlation coefficient (PCC) is a measure of linear correlation between two sets of data *X* and *Y*, which could be calculated as follows:

$$\mathrm{PCC} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}} \tag{14}$$

where $\bar{X}$ and $\bar{Y}$ are the average value of *X* and *Y*, respectively, and *n* is the size of *X* and *Y*.

Several authors have offered guidelines for the interpretation of a correlation coefficient (CC) [41]–[43].In these guidelines, |PCC| greater than 0.8 indicates a very strong correlation; |PCC| between 0.6 and 0.8 indicates a strong correlation;|PCC| between 0.4 and 0.6 indicates a moderate correlation; |PCC| between 0.2 and 0.4 indicates a weak correlation; and |PCC| less than 0.2 means there is no correlation.

In this study, to prove the two trends are correlated, PCC needed to be calculated. If the absolute value of PCC is over 0.2, it could be considered that there is at least a weak correlation between the two trends. Table III presents the PCC of the characteristic trends between each two years. If the trend of characteristics in each year is similar, it could be asserted it is reasonable to select these characteristics to establish the bidirectional associations.

To illustrate whether there are bidirectional associations between green tides and social responses, this study calculated the PCCs between the characteristics of Sina Weibo with the coverage of *U. prolifera* green tides. Table IV presents the PCCs between the characteristics of Sina Weibo and CA of green tides. If the characteristic selected in this study is correlated to the coverage of *U. prolifera* green tides, it could be asserted that there are bidirectional associations between the green tides and social response. A more detailed analysis is as follows.

*1) CA of 2016–2019 U.***Prolifera** *Green Tides:* In the existing literature, the development of *U. prolifera* green tides in

TABLE IV
PEARSON CORRELATION ANALYSIS BETWEEN CA AND CHARACTERISTICS OF SINA WEIBO IN EACH YEAR ($P<0.05$)

characteristic	Pearson correlation coefficients calculated between CA and characteristics of Sina Weibo			
	2016	2017	2018	2019
SDE_x	0.268574	-0.21885	0.371484	0.164314
SDE_y	0.668123	0.39711	0.640511	0.635414
σ_x	0.757731	0.416183	0.704179	0.661503
σ_y	0.440732	0.587878	0.311365	0.596667
α	0.773525	0.242164	0.773955	0.532527
$\sigma_{\mathrm{B,x}}(B = Beijing)$	0.498333	0.549922	0.387543	0.619864
$\sigma_{\mathrm{B},y}(B = Beijing)$	0.684778	0.645163	0.552473	0.706048
$\sigma_{\mathrm{B,x}}(B = Jinan)$	0.498333	0.549922	0.387543	0.619864
$\sigma_{\mathrm{B},y}(B = Jinan)$	0.684778	0.645163	0.552473	0.706048
$\sigma_{\mathrm{B,x}}(B = Qingdao)$	0.498333	0.549922	0.387543	0.619864
$\sigma_{\mathrm{B},y}(B = Qingdao)$	0.684778	0.645163	0.552473	0.706048
$\sigma_{\mathrm{B,x}}(B = Nantong)$	0.498333	0.549922	0.387543	0.619864
$\sigma_{\mathrm{B},y}(B = Nantong)$	0.684778	0.645163	0.552473	0.706048
$\rho(B)(B = Beijing)$	0.171837	-0.028237	0.265738	0.283409
$\rho(B)\ (B = Jinan)$	0.164972	-0.028062	0.259848	0.280134
$\rho(B)\ (B = Qingdao)$	0.166182	-0.031727	0.261423	0.278777
$\rho(B)\ (B = Nantong)$	0.25973	0.065565	0.310842	0.362358

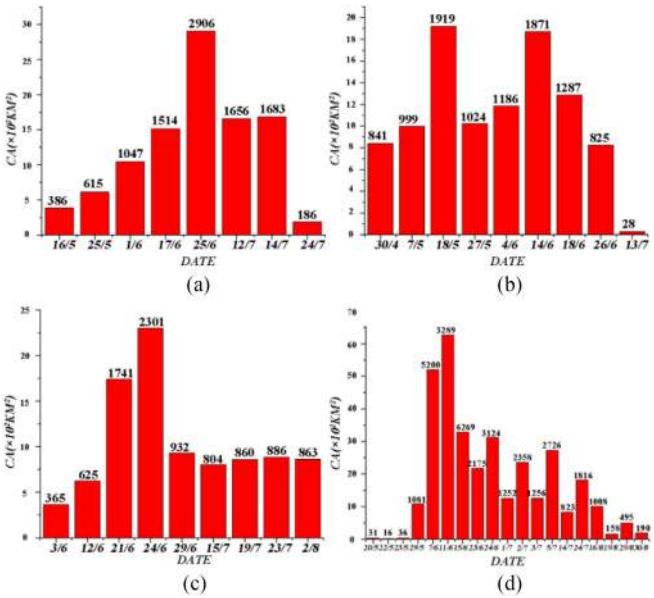


Fig. 11. CA of green tides in between 2016 and 2019. (a) 2016. (b) 2017. (c) 2018. (d) 2019.

the Yellow Sea was summarized as five major stages: appearance, development, outbreak, decline, and disappearance [2]. Fig. 11 illustrates CA on a different date in 2016–2019. In 2016 [Fig. 11(a)], 2018 [Fig. 11(c)], and 2019 [Fig. 11(d)], CA is consistent with the five stages, which appeared in mid-May to early June, developed in early June, peaked in mid-to-late June, then declined, and disappeared in late July to early August. In 2017 [Fig. 11(b)], the *U. prolifera* green tide appeared at the end of April, developed in early May. Then, the green tide peaked twice in mid-May to mid-June. After that, it shrunk quickly and almost died out in mid-July. In general, the trend of CA is following the five stages (appearance, development, outbreak, decline, and disappearance). In Table III, PCCs between the trend of CA in each year are all over 0.2, which means the trends of CA in 2016–2019 are similar.

*2) Gravity Centers of 2016–2019 U.***Prolifera** *Green Tides:* Fig. 12 illustrates the gravity centers $Q(\bar{x}, \bar{y})$ of *U. prolifera* green tides during 2016–2019. From the perspective of starting and endpoints of the drift paths of these four years, the *U. prolifera* green tide often originally grew on the southeast Yellow Sea and died out near Qingdao, Shandong Province. Among them, the green tide drift path in 2019 is the longest, which was due to the longest duration of 2019 *U. prolifera* green tides. The end date of the 2016–2018 green tides is from mid-July and early August before which the four years of disasters have similar characteristics. In Table III, PCCs between the trend of $Q(\bar{x}, \bar{y})$ in each year are all over 0.2, which means the trends of $Q(\bar{x}, \bar{y})$ in 2016–2019 are similar.

*3) Distribution of 2016–2019 U.***Prolifera** *Green Tides:* Fig. 13 illustrates the 2016–2019 *U. prolifera* green tide distribution direction θ. In this study, θ was divided according to the growth stage of *U. prolifera* (i.e., appearance, development, outbreak, decline, and disappearance). It is worth noting that the observed images of green tides are limited due to the clouds, the satellite coverage. Based on the limited data, it is difficult to identify the accurate start and end time of the five growth stages of green tides. However, the limited data used in this study could reflect the five growth stages of green tides. From Fig. 13, in the existence of green tides, there was an observed sudden change in θ during the period of development and outbreak, which is 74.5° in 2016, 128.9° in 2017, 163.7° in 2018, and 160.1° in 2019, respectively. The observed sudden changes in θ are due to the change of SSW direction [2]. After the observed sudden change is removed, the green tide of the same period presents a certain continuity in θ.

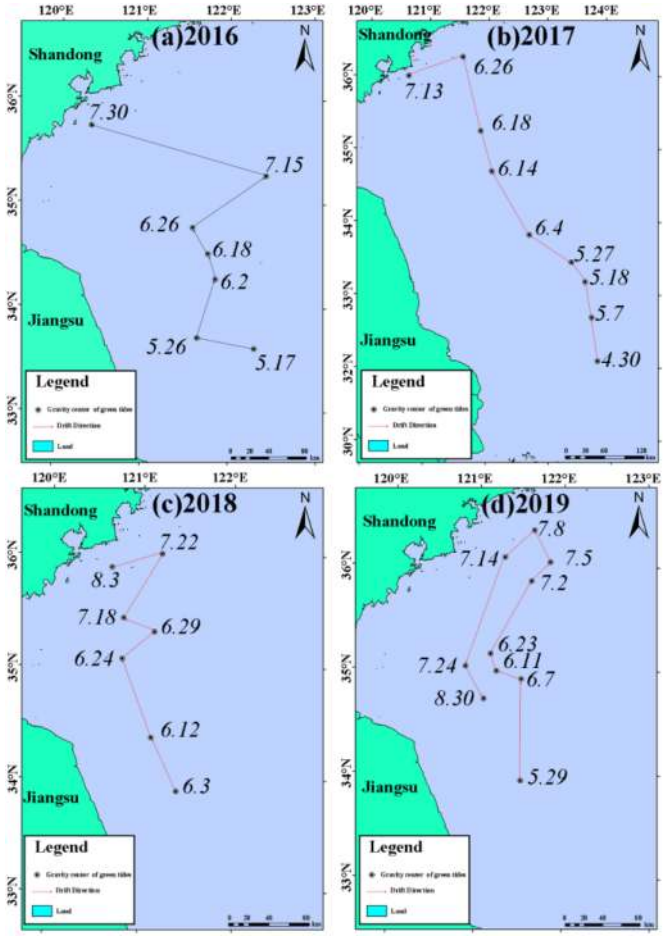


Fig. 12. Drift paths of green tides in between 2016 and 2019. (a) 2016. (b) 2017. (c) 2018. (d) 2019.

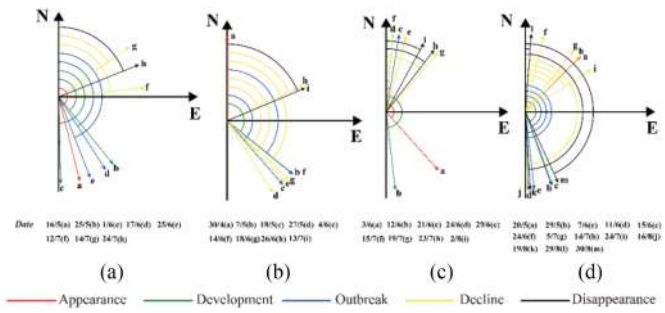


Fig. 13. θ of green tides in between 2016 and 2019. (a) 2016. (b) 2017. (c) 2018. (d) 2019.

Fig. 14 illustrates the distribution length L_1 and distribution width L_2 of the *U. prolifera* green tides during 2016–2019. From the perspective of L_1 and L_2, the green tides in these four years tend fluctuation increasing first and then decreasing, which conformed to the growth law of *U. prolifera*. The peak of L_1 was between 400 and 600 km, whereas the peak of L_2 was between 200 and 250 km. In Table III, PCCs between the trend of θ, L_1, and L_2 in each year are all over 0.2, which means the trends of θ, L_1, and L_2 in 2016–2019 are similar.

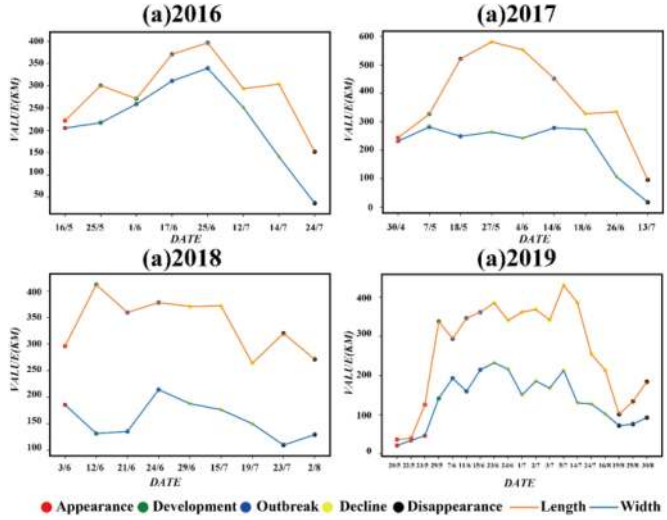


Fig. 14. L_1 and L_2 of green tides in between 2016 and 2019. (a) 2016. (b) 2017. (c) 2018. (d) 2019.

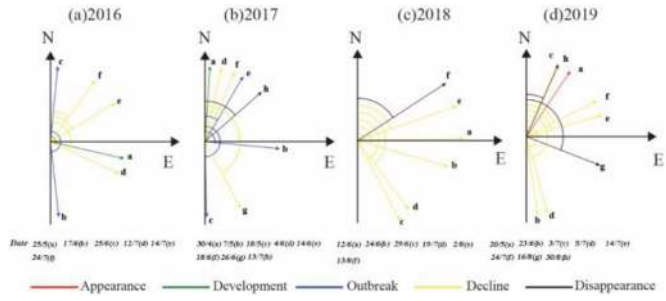


Fig. 15. α of Sina Weibo in between 2016 and 2019. (a) 2016. (b) 2017. (c) 2018. (d) 2019.

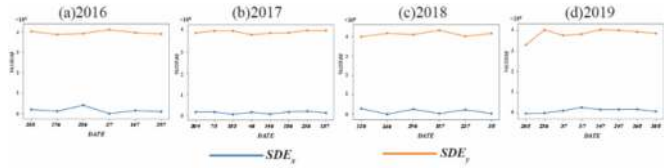


Fig. 16. SDE_x and SDE_y of Sina Weibo in between 2016 and 2019. (a) 2016. (b) 2017. (c) 2018. (d) 2019.

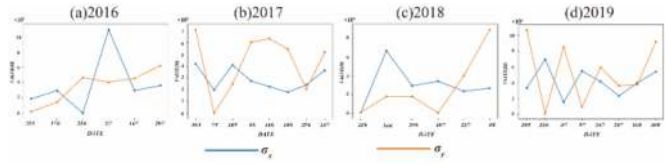


Fig. 17. σ_x and σ_y of Sina Weibo in between 2016 and 2019. (a) 2016. (b) 2017. (c) 2018. (d) 2019.

4) Morphological Characteristic Index of 2016–2019 Sina Weibo Data: Figs. 15 –17 show the α, $(\mathrm{SDE}_x, \mathrm{SDE}_y)$, and (σ_x, σ_y) of the Sina Weibo data. Due to the small amount of Sina Weibo data, α and (σ_x, σ_y) did not show a consistent pattern in these four years. However, $(\mathrm{SDE}_x, \mathrm{SDE}_y)$ did not change significantly from year to year. The SDE_x was always between

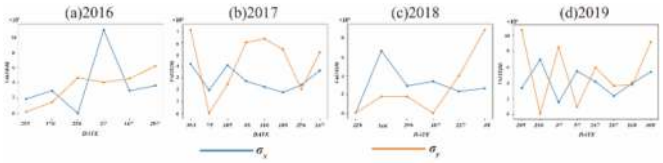


Fig. 18. $\rho(B)$ of Sina Weibo in between 2016 and 2019. (a) 2016. (b) 2017. (c) 2018. (d) 2019.

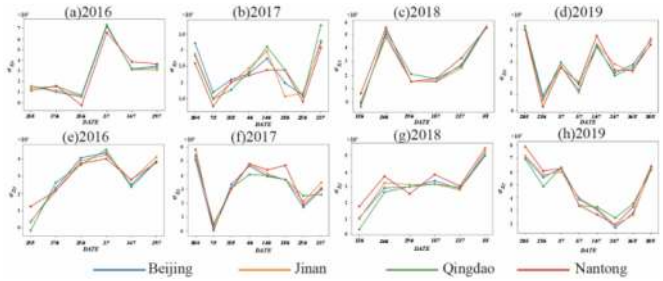


Fig. 19. $\sigma_{B,x}$ and $\sigma_{B,y}$ of Sina Weibo in between 2016 and 2019. (a) 2016. (b) 2017. (c) 2018. (d) 2019. (e) 2016. (f) 2017. (g) 2018. (h) 2019.

3×10^6m and 4×10^6m, whereas the SDE_y was always between 0 and 5×10^5m, which meant the annual center of Sina Weibo data was relatively consistent. Apart from the trend in 2017, PCCs between the trend of α, SDE_y, σ_x, and σ_y in each year are all over 0.2 in Table III, which means the trends of α, SDE_y, σ_x, and σ_y in each year are similar. In Table IV, the trend of CA with the trend of morphological characteristic index has a correlation (PCCs are over 0.2), which means it is reasonable to establish the bidirectional associations based on morphological characteristics.

5) Spatio-Temporal Point Patterns of 2016–2019 Sina Weibo Data: Fig. 18 illustrates the intensity function $\rho(B)$ of Sina Weibo data during 2016–2019. With the development of *U. prolifera* green tides, the intensity of Sina Weibo in Beijing, Jinan, Qingdao, and Nantong all showed an increasing trend. During the development of *U. prolifera* green tides, the intensity was always in order of Beijing, Jinan, Qingdao, and Nantong. Fig. 19 illustrates the $\sigma_{B,x}$ and $\sigma_{B,y}$ of Sina Weibo data during 2016–2019. In these four years, the changing trend of Beijing, Jinan, Qingdao, and Nantong were similar every year. Apart from the trend in 2017, PCCs between the trend of $\rho(B)$, $\sigma_{B,x}$, and $\sigma_{B,y}$ in each year are all over 0.2 in Table III, which means the trends of $\rho(B)$, $\sigma_{B,x}$, and $\sigma_{B,y}$ in each year are similar. In Table IV, the trend of CA with the trend of $\sigma_{B,x}$ and $\sigma_{B,y}$ has a correlation (PCCs are over 0.2), which means it is reasonable to establish the bidirectional associations based on $\sigma_{B,x}$ and $\sigma_{B,y}$. In 2018 and 2019, there are correlations between the trends of $\rho(B)$ and CA.

It is worth noting that the number of effective Weibo blogs used in this study is 1064.There are two reasons for accounting for that the Weibo data used in this study could be trusted. First, the characteristics extracted from Weibo data (i.e., $\mathrm{SDE_x}$, SDE_y, σ_x, σ_y, α, $\sigma_{B,x}$, $\sigma_{B,y}$) may have similar trends in each year. In Table III, apart from the trend in 2017, PCCs between the trend of α, SDE_y, σ_x, σ_y $\rho(B)$, $\sigma_{B,x}$, and $\sigma_{B,y}$ in each

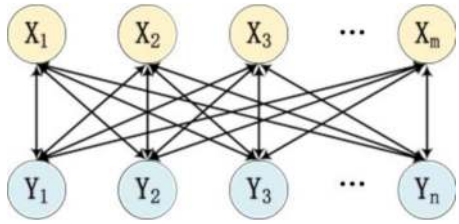


Fig. 20. Structure of the BAM.

two years are all over 0.2, which means the trends of α, SDE_y, σ_x, σ_y, $\rho(B)$, $\sigma_{B,x}$, and $\sigma_{B,y}$ in each year are similar. Based on the similar trends of the Sina Weibo, it could be concluded that the quality of Sina Weibo may be stable and may not change with the blog number. Second, if the number of Weibo data is large enough, the quality to describe the green tides of Weibo data will still be lower than that of RS images, which is due to the limited ability of Weibo data. In the designed model, the existing Weibo blogs were used to predict the MBR of green tides, whose quality is lower than the irregular polygon distribution of green tides

Based on the analysis above, this study asserted that the trend of characteristics in each year is similar and the characteristics of Sina Weibo and green tides are correlated, which means there are bidirectional associations between the green tides and social response. After illustrating the bidirectional association between green tides and Sina Weibo blogs, BSAMNN was designed to establish the bidirectional association. This study used the green tides and Sina Weibo data in 2016–2018 as training data to construct BSAMNN. The trend of green tides' and Sina Weibo's characteristics used in this study are similar in each year, so it is reasonable to use these characteristics to train BSAMNN. Based on the consistency and correlation between the times series of green tides and Sina Weibo blogs, the bidirectional association could be captured by BSAMNN. To test the feasibility of BSAMNN, characteristics of green tides and Sina Weibo data in 2019 are used as testing data.

D. Construction of BSAMNN

This section introduces the basic BAM approach first. Second, characteristics of Sina Weibo were modified with geographic weight. Third, the residual network and spatial-temporal constraints, which were used in this study, were defined. Fourth, the spatial-temporal constraints of SSW and ocean surface current were established. Finally proposes the construction of BSAMNN.

1) Bidirectional Associative Memory Neural Network: BAM is a two-layer Artificial Neural Networks ANN, whose two layers are named as *X*-layer and *Y*-layer. Fig. 20 illustrates the basic structure of BAM. The *X*-layer contains *m* neurons (i.e., CA, θ, L_1, L_2, and $Q(\bar{x}, \bar{y})$ of green tides) and the *Y*-layer contains *n* neurons (i.e., $\alpha, \sigma_x, \sigma_y, \rho(B), \sigma_{B,x}, \sigma_{B,y}$, emotion).

As is illustrated in Fig. 20, BAM is fully connected and each neuron in *X*-layer and *Y*-layer is connected to each neuron in the other layer. The neuron in BAM is a bipolar model whose output is either –1 or +1 and the activation function is signum

function. The updating rule of each neuron is as (15)–(18) [44]

$$y_j(t+1) = f\left(\sum_{i=1}^{m} W_{ij} x_i(t), y_j(t)\right) \tag{15}$$

$$x_j(t+1) = f\left(\sum_{i=1}^{m} W_{ij}^T y_i(t), x_j(t)\right) \tag{16}$$

$$W = \sum_{k=1}^{p} x_k y_k^T \tag{17}$$

$$f(a,b) = \begin{cases} 1, & a > 0 \\ b, & a = 0 \\ -1, & a < 0 \end{cases} \tag{18}$$

where t is the iteration number, $x_i(t)$ and $y_i(t)$ are the status of the neurons in *X*-layer and *Y*-layer, respectively, W is the $m \times n$ weight matrix, which can be calculated as (16), f is the neurons activation function, which is defined as (17), and p is the number of stored patterns.

As is shown in (14) and (15), when a given pattern is applied to BAM, the most similar stored pair of patterns will be received [44]. Depending on the input pattern, BAM adjusts the outputs of neurons in *X*-layer and *Y*-layer to +1 or –1. Then, the neurons in *X*-layer and *Y*-layer are iteratively updated based on their weights and their previous values. After several iterations, the outputs of the *X*-layer and *Y*-layer converged to the most similar storage mode.

This study viewed CA, θ, L_1, L_2, $Q(\bar{x}, \bar{y})$ of green tides as the patterns of *X*-layer and views α, σ_x, σ_y, $\rho(B)$, $\sigma_{B,x}$, $\sigma_{B,y}$, emotion and $(\mathrm{SDE}_x, \mathrm{SDE}_y)$ as the pattern of *Y*-layer. It is worth noting that the above variables are all continuous. This study treated each continuous variable as a 12-b binary form and treated each bit of variable under binary as a neuron.

2) Modify the Characteristics of Sina Weibo ***With*** *Geographic Weight:* The annual green tides in the Yellow Sea usually appears near Nantong in Jiangsu Province and disappears near Qingdao in Shandong Province [6], which means citizens in Nantong and Qingdao will first find occurrence of *U. prolifera* green tides and post blogs in social media. In this study, $\sigma_{B,x}$, $\sigma_{B,y}$, and $\rho(B)$ with $B =$ Qingdao and Nantong should have higher weight in the neural network. To modify the characteristics of Sina Weibo blogs with geographic weight, this study used entropy weights method (EWM) [45], [46]. $\sigma_{B,x}$, $\sigma_{B,y}$, and $\rho(B)$ will be brought into EWM for weight calculation, respectively.

The EWM used in this study can be described as follows.

Step 1: Normalization of the characteristics. Let $C(B)$ $(B = Beijing, Jinan, Qingdao, and\ Nantong)$ as one of $\sigma_{B,x}$, $\sigma_{B,y}$ and $\rho(B)$. Let C_1, C_2, C_3, C_4 as $C(Beijing)$, $C(Jinan)$, $C(Qingdao)$, $C(Nantong)$, respectively. Let *N* is the length of $C(B)$, and C_1, C_2, C_3, C_4 can be normalized as follows:

$$C'_{i,j} = \frac{C_{i,j} - \min(C_j)}{\max(C_j) - \min(C_j)} \tag{19}$$

where $i = 1, 2, \ldots, N$ and $j = 1, 2, 3, 4$.

TABLE V
ENTROPY WEIGHT OF $\sigma_{B,x}$, $\sigma_{B,y}$, AND $\rho(B)$

Hot Points	$\rho(B)$	σ_{B,x_n}	σ_{B,y_n}
Beijing	0.0131	0.0525	0.2499
Jinan	0.0375	0.0599	0.2499
Qingdao	0.7048	0.4498	0.2500
Nantong	0.2447	0.4378	0.2500

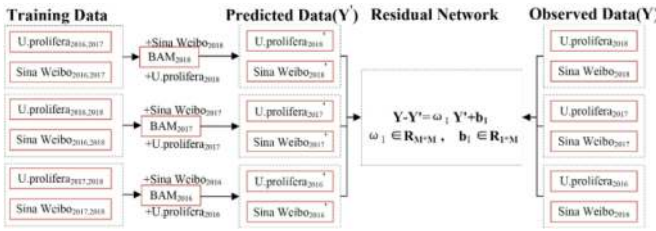


Fig. 21. Training process of the residual network.

Step 2: Probability and entropy calculation. The probability $p_{i,j}$ can be calculated as (20) and the entropy of E_j can be calculated as (21)

$$p_{i,j} = \frac{C'_{i,j}}{\sum_{i=1}^{N} C'_{i,j}} \tag{20}$$

$$E_j = -\frac{1}{\ln N} \sum_{i=1}^{N} p_{i,j} \ln p_{i,j}. \tag{21}$$

Step 3: Weight calculation. After calculating E_j, the entropy weight(EW_j) of C_j can be obtained as follows:

$$EW_j = \frac{1 - E_j}{4 - \sum_{j-1}^{4} E_j}. \tag{22}$$

The EW_j of $\sigma_{B,x}$, $\sigma_{B,y}$, and $\rho(B)$ are presented in Table V. $\sigma_{B,x}$, $\sigma_{B,y}$, and $\rho(B)$ will be modified by the following equation:

$$C_j \leftarrow EW_j * C_j. \tag{23}$$

3) Construction of the Residual Network: Fig. 21 illustrates the training process of the residual network. This study first constructed the dataset based on the green tides and Sina Weibo data of 2016, 2017, and 2018. The datasets of 2016, 2017, and 2018 were set as verification dataset in turn and the corresponding remaining data as training data. Then, three sets of training data are separately trained in BAM and BAM_{2018}, BAM_{2017}, and BAM_{2016} were obtained. Finally, based on the characteristics of green tides and Sina Weibo, respectively, BAM_{2018}, BAM_{2017}, and BAM_{2016} were used to obtain the BAM results from 2016 to 2018.

To correlate the BAM results and the observed data, this study learned a residual network [47] considering the similarity between them. Denote the characteristics of *U. prolifera* green tides and Sina Weibo data during 2016–2018 as *Y*, the characteristics of BAM results during 2016–2018 as Y'. Then, the residual

network can be defined as follows:

$$Y - \ Y' = \omega_1 \ Y' + b_1 \tag{24}$$

where ω_1 and b_1 are the weight and bias of the residual network, respectively.

4) Construction of the Spatio-Temporal Constraints Through Characteristics: This study asserted that the annual *U. prolifera* green tides and its social media data have a bidirectional association. Thus, this study uses the BAM and residual network to explore these associations between them. However, the annual *U. prolifera* green tides and its social media data have their spatio-temporal constraints, which may be violated by the BAM results. For instance, after the occurrence of green tides, its distribution gravity center is most likely moving north [2]. To obtain the spatio-temporal constraints of the *U. prolifera* green tides and Sina Weibo data, respectively, this study used the association rule mining, which is a popular data mining method for discovering association relations among various variables from massive data [48]. "*A*→*B*" is the most common form of an association rule, where *A* and *B* are called the antecedent and the consequent, respectively. The typical indicators to measure the strength of an association rule are support and confidence [49]. Let *D* denotes the characteristic dataset of *U. prolifera* green tides or Sina Weibo data. The $support$ of a rule is the percentage of transactions in *D* contain both *A* and *B*, which is often used to describe how often *A* and *B* co-occur in *D* and can be defined as (25). The confidence of a rule is used to estimate of the conditional probability *P*(*B*|*A*) and can be defined as (26)

$$\text{support} \ \ (A \to B) = P\left(d \in D; A, B \subseteq d\right) \tag{25}$$

$$\text{confidence} \ \ (A \to B) = P \ \ (BA) = \frac{P\left(d \in D; A, B \subseteq d\right)}{P\left(d \in D; A \subseteq d\right)}. \tag{26}$$

This study compared the predicted Y', which obtained only by BAM with the observed data Y. The average absolute difference (AAD), the average relative difference (ARD), and CC between the predicted and observed data were used to evaluate the performance of BAM results. The definitions of AAD, ARD, and CC are as (27)–(29)

$$\text{AAD} = \frac{\sum_{i=1}^{N} |Y_i - {Y'}_i|}{N} \tag{27}$$

$$\text{ARD} = \frac{\sum_{i=1}^{N} \left|\frac{Y_i - {Y'}_i}{Y_i}\right|}{N} \tag{28}$$

$$\text{CC} = \frac{\sum_{i=1}^{N} \left[\left(Y_i - \bar{Y}\right)\left({Y'}_i - \overline{Y'}\right)\right]}{\sqrt{\sum_{i=1}^{N} \left(Y_i - \bar{Y}\right)^2} \sqrt{\sum_{i=1}^{N} \left({Y'}_i - \overline{Y'}\right)^2}}. \tag{29}$$

AAD shows the average residual, ARD shows the percentage of the residuals relative to the observed data, and CC shows the linear relationship between the predicted data and observed data. This study selected the characteristics obtained by BAM with better performance in terms of AAD, ARD, and CC as the antecedent *A* in the association rule mining.

Fig. 22 shows the difference between the results in 2019 directly from BAM and the observed data. Among the characteristics of *U. prolifera* green tide, CA and θ achieved the best performance in AAD, ARD, and CC. The AAD of CA and θ were 226.29 and 9.71, respectively, which was far less than other characteristics. It is worth noting that the AAD of L_1, L_2, $Q(\bar{x}, \bar{y})$ was over 10^5 because they were related to the coordinate system. In terms of CC, the linear relationship between the predicted data and observed data of CA and θ were all over 0.85 and just less than L_2. In terms of ARD, the percentage of residuals relative to the observed data of CA and θ were 0.11 and 0.13. Therefore, this study selected CA and θ as A in the *U. prolifera* green tide association rule mining.

Among the characteristics of Sina Weibo data, $\rho(B)(B = Beijing, Jinan, Qingdao, \ and \ Nantong)$ achieved the best performance. The CC of them was 0.99, 1.00, 1.00, and 0.99, respectively, which means there was a linear correlation between the predicted and observed data. The ARD of them is 0.10, 0.06, 0.08, and 0.24, which means the residuals of them are relative to the observed data. The AAD of them was 8.31, 2.15, 3.15, and 2.15, which was smaller than other characteristics with a similar CC and ARD. Therefore, this study selected CA and θ as A in the Sina Weibo data association rule mining.

Fig. 23 gives an overview of the calculation flow used in obtaining the spatio-temporal constraints in this study. The typical association rule mining methods (i.e., FP-Growth and Apriori) extract rules from discrete data, whereas the characteristics used in this study were all continuous data. Therefore, the calculation flow first discretizes these characteristics and extracts rules from these discrete data. The calculation flow is based on the genetic algorithm and Apriori algorithm and can be described as follows.

Step 1: Initializing the population pop_1. Here, pop_1 is used to describe how many discrete intervals that each characteristic needs to be formed. An individual in pop_1 represents a scheme for the number of discreating the characteristics. The *i*th chromosome in an individual represents the *i*th characteristic's cluster number. Each gene corresponds to the one of binary forms of a chromosome. An individual in pop_1 represents a combination of the cluster number of all characteristics.

Step 2: Initializing the population pop_2. Here, pop_2 is used to define the cluster centers of each characteristic. The number of each characteristic's cluster centers is corresponding to pop_1.An individual in pop_2 represents a scheme for the cluster center of each characteristics. The gene in pop2 denotes one of cluster center of a characteristic, the chromosome in pop_2 denotes one of the characteristics and an individual represents a discrete scheme of all characteristics. In

Steps 3 and 5: Evaluate the pop_1 and pop_2. This study used Apriori algorithm to extract the rules from pop_2. Then, this study used the sum of rule confidence, which is shown in (24) to evaluate the pop_2. When the termination condition is satisfied, that is the pop_2 generation is over the designed number of generation, the algorithm will submit the rules, which corresponding to the individuals of pop_1, to Step 5. Step 5 evaluates pop_1 via (30). When the termination

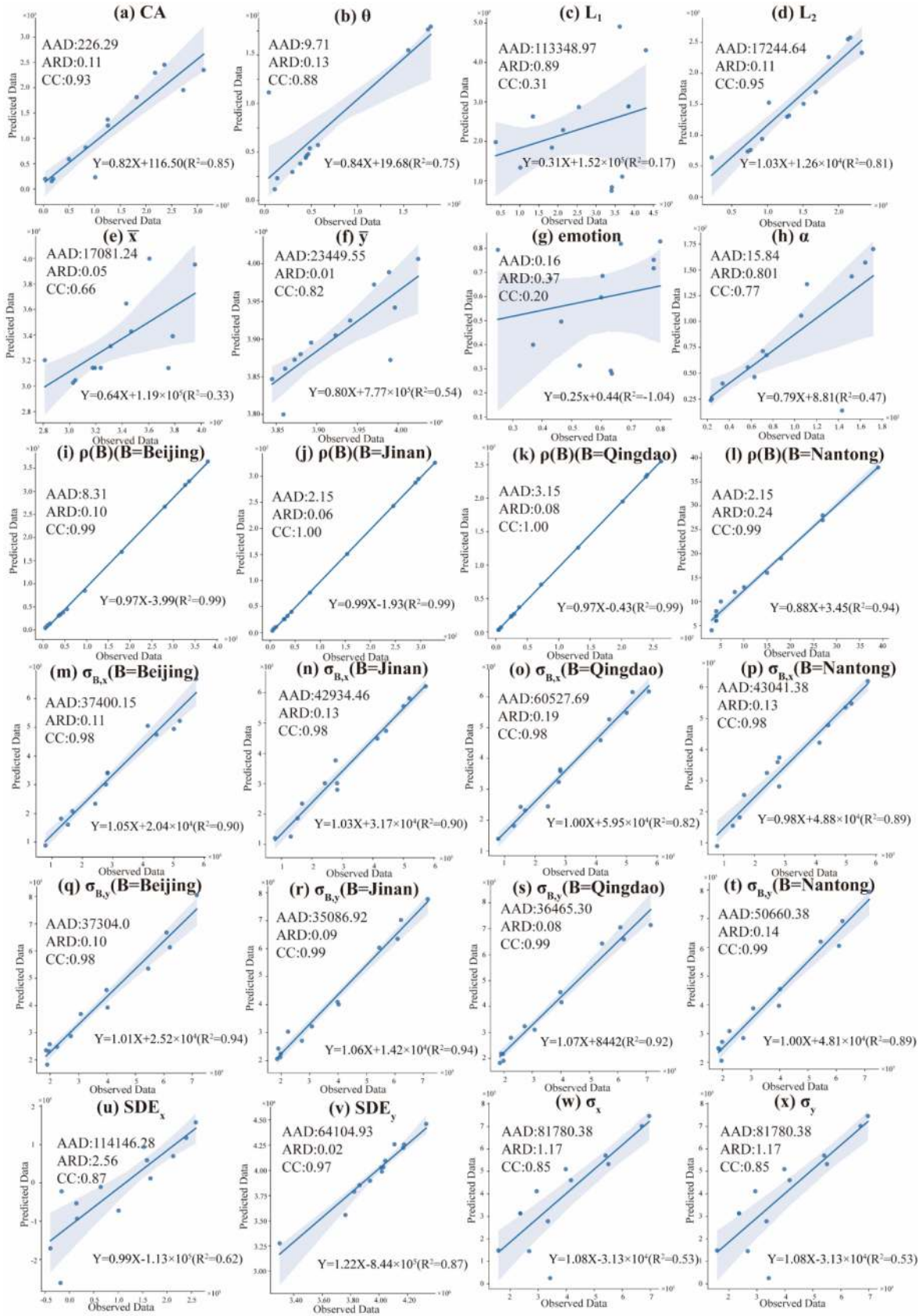


Fig. 22. Comparisons between the predicted results from BAM with the observed data.

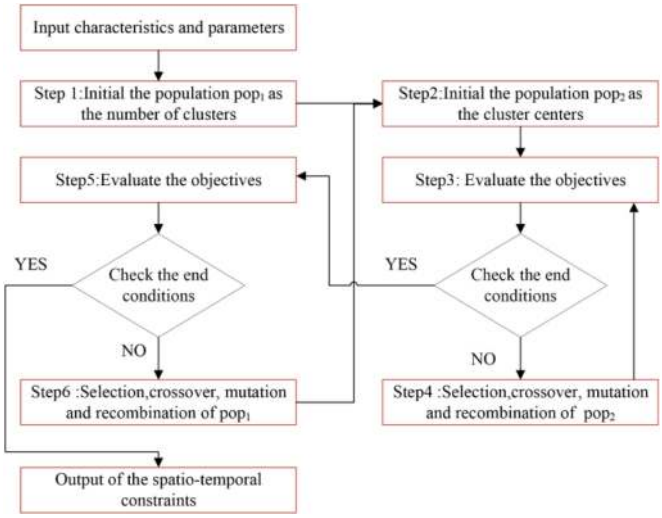


Fig. 23. Flowchart of the spatio-temporal constraints.

condition is satisfied, that is pop_1 generation is over the designed number of pop_1 generation, the algorithm will output the rules as the spatio-temporal constraints. Let denote the C as the constraints.

Steps 4 and 6: Selection, crossover, mutation, and recombination of pop_1 and pop_2. The selection, crossover, mutation, and recombination of pop_1 and pop_2 are the same as the genetic algorithm. The detailed explanation of them can be found in the work of Molahasani Majdabadi *et al.* [44]

$$\max F = \sum_{i}^{\text{rule}} \text{confidence}\left(A \to B\right). \tag{30}$$

5) Construction of the Spatio-Temporal Constraints Through SSW and Ocean Surface Current: This study proved that there are bidirectional associations between *U. prolifera* green tides in the Yellow Sea and Sina Weibo data. However, the *U. prolifera* green tides in the Yellow Sea are mainly affected by SSW and ocean surface current. In the work of Cao *et al.* [18], the drift paths of green tides in the Yellow Sea are consistent with the SSW direction. So, the gravity center $Q(x, y)$ of green tides needs to consistent with the SSW direction and ocean surface current. When there have been green tides detected in this month, the constraints, which are shown in (31), will be added to the spatio-temporal constraints

$$\left\lfloor \frac{\text{DIR}}{90} \right\rfloor \times 90 \leq \arctan \left| \frac{y - y_0}{x - x_0} \right| \leq \left\lfloor \frac{\text{DIR}}{90} + 1 \right\rfloor \times 90 \tag{31}$$

where (x_0, y_0) is the last gravity center of green tides in this month, and DIR is the average SSW direction or ocean surface current direction in this month.

E. Construction of BSAMNN

In this study, residual networks and spatio-temporal constraints were used to improve the results of BAM. The improved bidirectional associative memory network is described as (32) and (33). In the process of BSAMNN operation, if the results meet the spatio-temporal constraints, the results are considered as the outputs that make the results more in line with the characteristics of green tides and Sina Weibo data themselves. When BSAMNN runs to a stable state, the residual networks will be used to correct the outputs, which can help reduce the BAM errors

$$y_j\left(t+1\right) = \begin{cases} \text{sgn}\left[\sum\limits_{i=1}^{m} W_{ij} x_i\left(t\right)\right], \sum\limits_{i=1}^{m} W_{ij} x_i\left(t\right) \neq 0 \text{ and} \\ \quad y_j\left(t+1\right) \notin C \\ \omega_1\left(y_j\left(t\right) + I\right) + b_1, \sum\limits_{i=1}^{m} W_{ij} x_i\left(t\right) = 0 \\ y_j\left(t\right), \sum\limits_{i=1}^{m} W_{ij} x_i\left(t\right) \neq 0 \text{ and } y_j\left(t+1\right) \in C \end{cases} \tag{32}$$

$$x_j\left(t+2\right) = \begin{cases} \text{sgn}\left[\sum\limits_{i=1}^{m} W_{ij}^T y_i\left(t+1\right)\right], \sum\limits_{i=1}^{m} W_{ij}^T y_i\left(t+1\right) \neq 0 \\ \quad \text{and } x_j\left(t+2\right) \notin C \\ \omega_1(x_j\left(t\right) + I) + b_1, \sum\limits_{i=1}^{m} W_{ij}^T y_i\left(t+1\right) = 0 \\ x_j\left(t+1\right), \sum\limits_{i=1}^{m} W_{ij}^T y_i\left(t+1\right) \neq 0 \\ \quad \text{and } y_j\left(t+2\right) \in C \end{cases} \tag{33}$$

IV. Experiment and Results

A. Experiment Results From BSAMNN

This study used the green tides and Sina Weibo data in 2016–2018 as the training dataset of BSAMNN. To test whether BSAMNN could be used in the detection of green tide detection, dataset in 2019 were used as testing data. The number of Sina Weibo blogs in training and testing data is 624 and 440, respectively. The number of green tide detection result in training and testing data is 26 and 13, respectively.

1) Visualization of the BSAMNN and BAM Results: Fig. 24 shows the visualization of the BSAMNN and BAM results in 2019. In Fig. 21, this study found that BAM could capture the gravity center of green tides during the development, appearance, and decline of *U. prolifera*. However, the MBR from BAM is smaller than the observed MBR, which means that there are serious missing alarms in BAM results. During the appearance and disappearance of *U. prolifera* green tides, the MBR is larger than the observed, which means there are serious false alarms. Compared with BAM, BSAMNN improved the results by reducing the false alarms when *U. prolifera* appears and disappears and reducing the missing alarms when *U. prolifera* develops, outbreaks, and declines. The quantitative analysis will be described in the following sections.

2) Assessment of BSAMNN: Fig. 25 shows the difference between the results in 2019 from BSAMNN and the observed data. AAD, ARD, CC will be used to evaluate the results. In Fig. 22, this study concluded that CA, θ of *U. prolifera* green tides and $\rho(B)(B = Beijing, Jinan, Qingdao, Nantong)$ from BAM results achieved the best performance in AAD, ARD, and CC. The CA of BSAMNN results achieved 228.90, 0.35, and 0.94, respectively, in AAD, ARD, and CC, whereas θ achieved 18.94, 0.59, and 0.91, respectively, which meant that BSAMNN

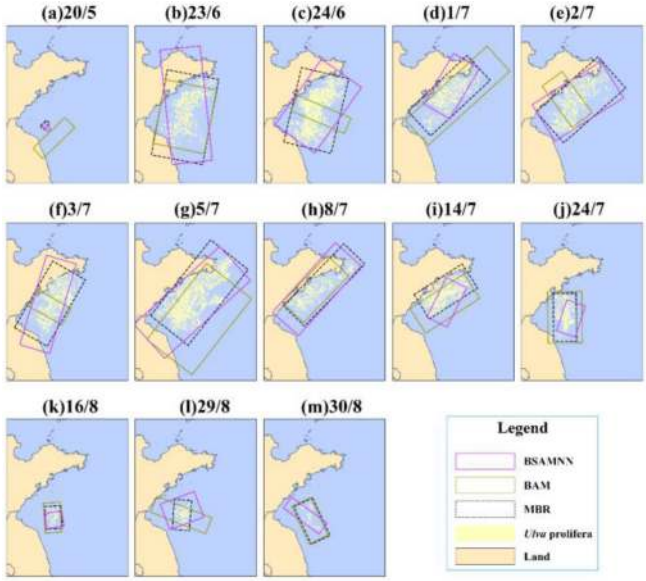


Fig. 24. Visualization of BSAMNN and BAM results in 2019.

could predict CA and θ with more confidence. In terms of $\rho(B)$, the ARD was less than and equal to 0.01, and CC was greater than and equal to 0.99, which meant $\rho(B)$ of BSAMNN results was similar to the observed data.

Besides the six characteristics of BSAMNN above, L_1, $\bar{x}$ and $\bar{y}$ of *U. prolifera* green tides achieved better performance than BAM results. The ARD of L_1, $\bar{x}$, and $\bar{y}$ went down by 62351.17, 3305.39, and 1146.06. The CC of L_1, $\bar{x}$ and $\bar{y}$ went up by 0.62, 0.17, and 0.06. The ARD of L_1 went down by 0.60. These improvements indicated that BSAMNN improved the accuracy of green tide prediction compared with BAM.

$\sigma_{B,x}, \sigma_{B,y}$ ($B =$ Beijing, Jinan, Qingdao, Nantong) were two characteristics with significant improvements by BSAMNN. The average AAD of $\sigma_{B,x}$ went down on average 20 914.5, whereas that of $\sigma_{B,y}$ went down by 13 896, which meant that BSAMNN can simulate Sina Weibo data via the Poisson group. Concerning the SDE (i.e., α, SDE_x, SDE_y, σ_x, and σ_y), there were no obvious improvements compared with BAM, which was due to that there is no periodicity in α, σ_x, and σ_y and BAM has high accuracy in SDE_x and SDE_y.

B. Accuracy, FAR, and MAR

To further illustrate the improvement of BSAMNN on the accuracy of green tide detection compared with BAM, the accuracy, FAR, and MAR were used to evaluate the experimental results. Accuracy is used to describe the probability that accurate predictions. FAR is used to describe the probability that green tides do not exist but are predicted to exist. MAR is used to describe the probability that green tides exist but are predicted to be nonexistent. Accuracy, FAR, and MAR can be calculated as (34), (35), and (36), respectively

$$\mathrm{Accuracy} = \frac{S_{\mathrm{MBR}_p \cap \mathrm{MBR}_o}}{S_{\mathrm{MBR}_p}} \tag{34}$$

TABLE VI
STATISTICAL SIGNIFICANCE TESTING OF THE RESULTS

Green tide characteristic	P-value	Sina Weibo characteristic	P-value
CA	0.4386	α	0.6586
L_1	0.3446	$\sigma_{\mathrm{B,x}}$	all less than 0.1
L_2	0.0198	$\sigma_{\mathrm{B},y}$	all less than 0.1
$\bar{x}$	0.8190	$\rho(B)$	all less than 0.05
$\bar{y}$	0.0740	SDE_x	0.4199
θ	0.4615	SDE_y	0.2740
FAR	0.3264	σ_x	0.5498
MAR	0.3445	σ_y	0.5092
Accuracy	0.3264	emotion	0.4381

$$\mathrm{FAR} = \frac{S_{\mathrm{MBR}_p} - S_{\mathrm{MBR}_p \cap \mathrm{MBR}_o}}{S_{\mathrm{MBR}_p}} \tag{35}$$

$$\mathrm{MAR} = \frac{S_{\mathrm{MBR}_o} - S_{\mathrm{MBR}_p \cap \mathrm{MBR}_o}}{S_{\mathrm{MBR}_o}} \tag{36}$$

where MBR_p and MBR_o denotes the MBR of predicted data and observed data, respectively.

Fig. 26 illustrates the accuracy, FAR, MAR of BSAMNN, and BAM in green tide detection. The results of BSAMNN achieved better performance than BAM, especially in the early stage of *U. prolifera* green tides. The average accuracy, FAR, and MAR of BSAMNN results were 0.69, 0.25, and 0.31, respectively, which was 0.11 higher, 0.10 lower, and 0.11 lower than that of BAM results, which meant that BSAMNN could help SOA map the *U. prolifera* green tides daily in the absence of RS images.

From the perspective of accuracy [see Fig. 26(a)], BSAMNN was superior to BAM in the early stage of *U. prolifera* green tides (i.e., appearance, development, and outbreak), with an average of 32.42% higher. However, the accuracy of BSAMNN was lower 8.56% than that of BAM in the stage of decline and disappearance. It could be concluded that BSAMNN is helpful for the early estimation of green tides.

From the perspective of FAR [see Fig. 26(b)], FAR of BSAMNN was reduced by an average of 8.38% compared with that of BAM. It is worth noting that FAR is large on the 20th of May ($\mathrm{FAR}_{\mathrm{BAM}} = 1.0$ and $\mathrm{FAR}_{\mathrm{IBAM}} = 0.52$) and 29th August ($\mathrm{FAR}_{\mathrm{BAM}} = 0.70$ and $\mathrm{FAR}_{\mathrm{IBAM}} = 0.56$), which was due to that both BAM and BSAMNN made wrong predictions about the gravity center of *U. prolifera* green tides.

From the perspective of MAR [see Fig. 26(c)], MAR of BSAMNN was 10.35% lower on average than that of BAM, which was consistent with the drop of FAR. It is worth noting that the average MAR of BAM was 0.62 before 5th July, whereas that of BSAMNN was 0.29, which meant BSAMNN has a lower MAR in the early stage of green tide detection.

In order to illustrate whether BSAMNN results were significantly better than the BAM results, this study used a two-sample one-side *T*-test to evaluate the experimental results. Table VI

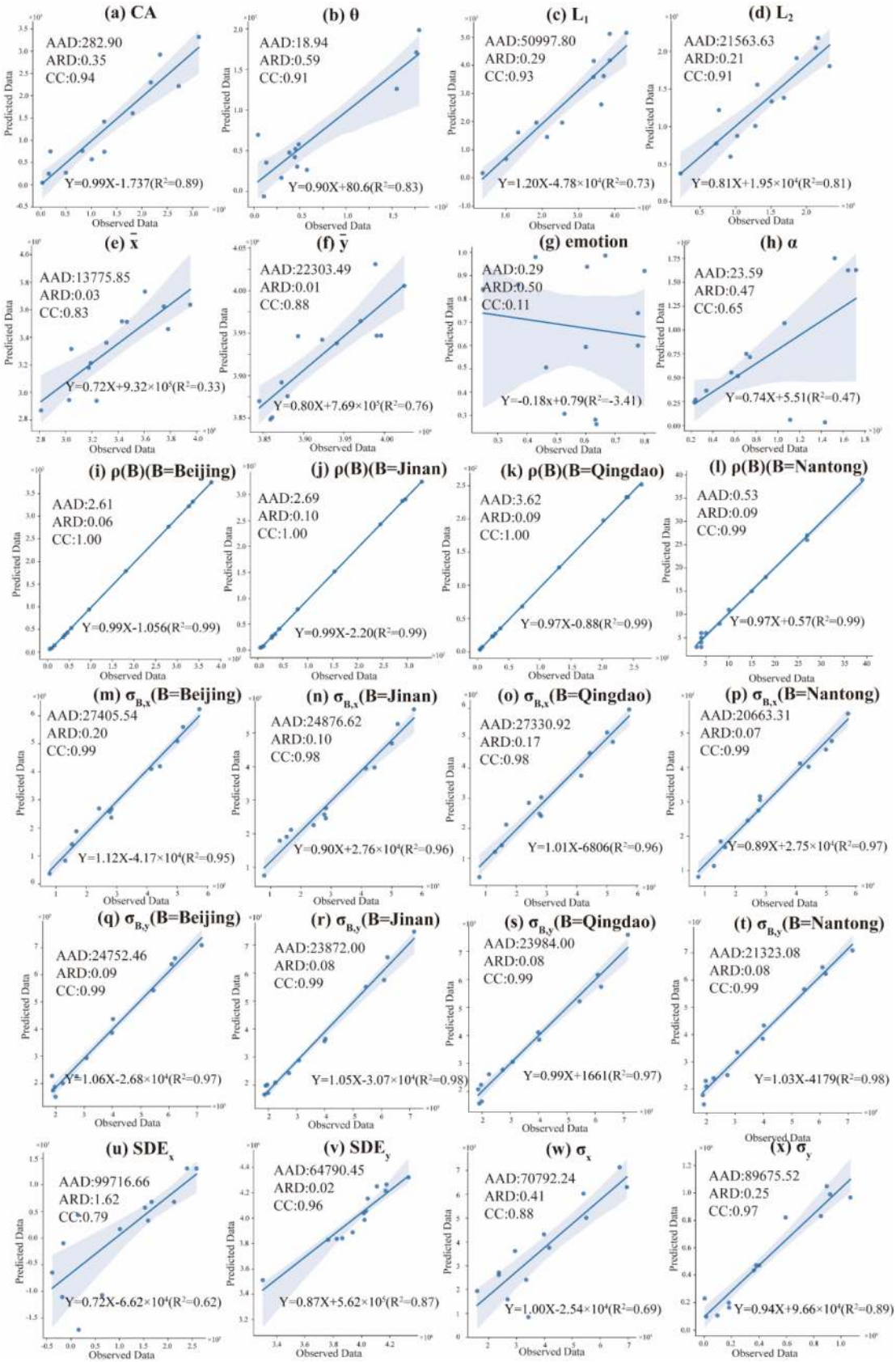


Fig. 25. Comparison between the observed data and the BSAMNN results.

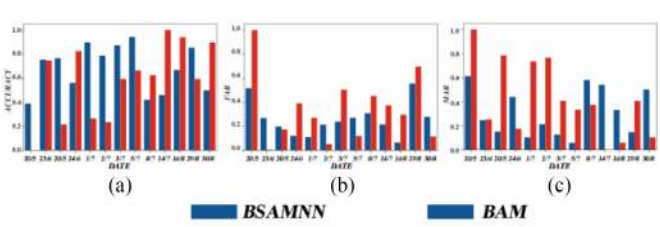


Fig. 26. (a) Accuracy, (b) FAR, and (c) MAR of the BSAMNN and BAM results.

TABLE VII
STATISTICAL SIGNIFICANCE TESTING OF THE RESULTS (CA>1000 KM^2)

Green tide characteristic	P-value	Sina Weibo characteristic	P-value
CA	0.5142	α	0.8957
L_1	0.1000	$\sigma_{B,x}$	all less than 0.1
L_2	0.0063	$\sigma_{B,y}$	all less than 0.1
$\bar{x}$	0.8470	$\rho(B)$	all less than 0.05
$\bar{y}$	0.0827	SDE_x	0.2769
θ	0.3251	SDE_y	0.4799
FAR	0.0722	σ_x	0.7335
MAR	0.4937	σ_y	0.8545
Accuracy	0.0722	emotion	0.6809

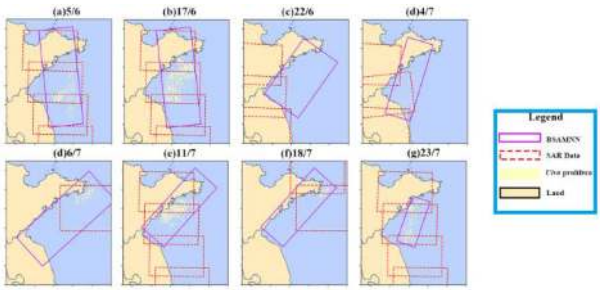


Fig. 27. BSAMNN results and the results from SAR images.

presents the statistical significance testing of the results. Table VII presents the statistical significance testing of the results when CA>1000 km^2.

Compared with BAM, the significantly improved variables of BSAMNN were L_1, $\bar{y}$, $\sigma_{B,x}$, $\sigma_{B,y}$, and $\rho(B)$(P<0.1), which caused the results of BSAMNN are better than that of BAM. When CA>1000 km^2, the accuracy and FAR of BSAMNN were significantly better than that of BAM(P<0.1), which meant BSAMNN had a better performance in mapping the *U. prolifera* green tides with a larger CA.

C. Comparing the Experiment Results With Those From SAR Images

Limited by the coverage and repeat circle of SAR images, green tides were found on June 5, June 17, June 22, July 4, July 6, July 11, July 18, and July 23 on SAR images. Fig. 27 compares the BSAMNN results and the results from SAR images on these

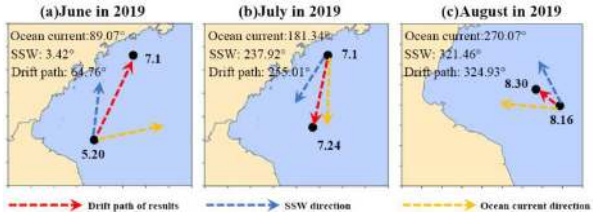


Fig. 28. Comparing the drift path from BSAMNN with SSW and ocean surface current direction.

eight days. In Fig. 25, Sentinel-1 GRD products covered only a part of the Yellow Sea in a day, which meant Sentinel-1 GRD products could not meet the requirements of daily mapping of *U. prolifera* green tides and the *U. prolifera* green tides shown in Fig. 27 are not completed. In these eight days, the accuracy of BSAMMN results was all over 97.27% and the mean accuracy is 98.97%, which meant the BSAMNN can cover the most of results obtained from SAR images.

D. Comparing the Experimental Results With SSW and Ocean Current Direction

The drift path of *U. prolifera* green tides in the Yellow Sea is consistent with the SSW direction [2], [28]. Fig. 28 shows the drift path of the experiment result and the Monthly SSW direction. In 2019, the SSW direction is 3.42°, 237.92°, 321.46° in June, July, and August. The ocean surface current direction is 64.76°, 255.01°, 324.93° in June, July, and August. Corresponding to SSW and ocean surface current, the direction of the green tide drift path is 3.42°, 255.01°, 324.93° in June, July, and August. In June, green tides in BSAMNN moved from southwest to northeast. In July, green tides in BSAMNN moved from the north to the south. In August, green tides in BSAMNN moved from southeast to northwest. These drift paths of experiment results are similar to the monthly SSW direction and ocean surface current direction in the three months. Fig. 28 illustrates that the drift path direction of experiment results is consistent with the direction of SSW and ocean surface currents.

V. CONCLUSION

This article first illustrated that there are bidirectional associations between green tides and Sina Weibo data through the CA, MBR, gravity center of green tides and SDE, hot points and emotion as characteristics of Sina Weibo data. Then, this article introduced an approach, namely, BSAMNN, which supported the bidirectional spatio-temporal association between the daily observation results of *U. prolifera* green tides in the Yellow Sea and the social response in social media data. In this approach, several spatial indicators and patterns of *U. prolifera* green tides and Sina Weibo data as the neural of the network, which could be used to estimate the *U. prolifera* green tides and Sina Weibo data. The results of BSAMNN achieved better performance than BAM, especially in the early stage of *U. prolifera* green tides. The average accuracy, FAR, and MAR of BSAMNN results were 0.69, 0.25, and 0.31, respectively, which was 0.11 higher, 0.10 lower, and 0.11 lower than that of BAM results. The proposed

method could help SOA map the *U. prolifera* green tides daily in the absence of RS data, and infer the possible situation of public sentiments based on observed *U. prolifera* results. This approach provided a broader view to make full use of social media data for the daily detection and mapping of *U. prolifera* green tides, and provide a practical decision-making tool for guiding public sentiments within social media.

Future research works should consider the geographical registration of green tides text in Sina Weibo data. Apart from the geo-location of users supplied from Sina Weibo, many texts of Sina Weibo contain the geo-location of *U. prolifera* green tides in the Yellow Sea, which could verify *U. prolifera* green tide mapping. Future developments should be directed toward extracting the geo-location of *U. prolifera* and merging the BSAMNN results and the information. Ultimately, the goal was to obtain more accurate information about the *U. prolifera* green tide distribution in the Yellow Sea.

REFERENCES

[1] X. Liu, Z. Wang, and X. Zhang, "A review of the green tides in the Yellow Sea, China," *Mar. Environ. Res.*, vol. 119, pp. 189–196, Aug. 2016, doi: 10.1016/j.marenvres.2016.06.004.

[2] Y. Cao, Y. Wu, Z. Fang, X. Cui, J. Liang, and X. Song, "Spatiotemporal patterns and morphological characteristics of Ulva prolifera distribution in the Yellow Sea, China in 2016–2018," *Remote Sens.*, vol. 11, no. 4, 2019, Art. no. 445, doi: 10.3390/rs11040445.

[3] Z. Wang, Z. Fang, Y. Wu, J. Liang, and X. Song, "Multi-Source evidence data fusion approach to detect daily distribution and coverage of Ulva prolifera in the Yellow Sea, China," *IEEE Access*, vol. 7, pp. 115214–115228, 2019.

[4] Y. Zhang *et al.*, "Ulva prolifera green-tide outbreaks and their environmental impact in the Yellow Sea, China," *Nat. Sci. Rev.*, vol. 6, no. 4, pp. 825–838, Jul. 2019, doi: 10.1093/nsr/nwz026.

[5] T. W. Cui *et al.*, "Assessing and refining the satellite-derived massive green macro-algal coverage in the Yellow Sea with high resolution images," *ISPRS J. Photogramm. Remote Sens.*, vol. 144, pp. 315–324, 2018, doi: 10.1016/j.isprsjprs.2018.08.001.

[6] P. Hu, Y. Liu, Y. Hou, and Y. Yin, "An early forecasting method for the drift path of green tides: A case study in the Yellow Sea, China," *Int. J. Appl. Earth Observ. GeoInf.*, vol. 71, pp. 121–131, Sep. 2018, doi: 10.1016/j.jag.2018.05.001.

[7] Y. Yang, J. Boncoeur, S. Liu, and P. Nyvall-Collen, "Economic assessment and environmental management of green tides in the Chinese Yellow Sea," *Ocean Coastal Manage.*, vol. 161, pp. 20–30, Jul. 2018, doi: 10.1016/j.ocecoaman.2018.04.012.

[8] C. Hu and M.-X. He, "Origin and offshore extent of floating algae in Olympic sailing area," *Eos Trans. AGU*, vol. 89, no. 33, pp. 302–303, Aug. 2008, doi: 10.1029/2008EO330002.

[9] C. Hu, "A novel ocean color index to detect floating algae in the global oceans," *Remote Sens. Environ.*, vol. 113, no. 10, pp. 2118–2129, Oct. 2009, doi: 10.1016/j.rse.2009.05.012.

[10] D. Liu, J. K. Keesing, Q. Xing, and P. Shi, "World's largest macroalgal bloom caused by expansion of seaweed aquaculture in China," *Mar. Pollut. Bull.*, vol. 58, no. 6, pp. 888–895, Jun. 2009, doi: 10.1016/j.marpolbul.2009.01.013.

[11] X. Sun *et al.*, "Spatio-temporal patterns of Ulva prolifera blooms and the corresponding influence on chlorophyll-a concentration in the southern Yellow Sea, China," *Sci. Total Environ.*, vol. 640/641, pp. 807–820, Nov. 2018, doi: 10.1016/j.scitotenv.2018.05.378.

[12] A. Ciappa, L. Pietranera, A. Coletta, and X. Jiang, "Surface transport detected by pairs of COSMO-SkyMed ScanSAR images in the Qingdao region (Yellow Sea) during a macro-algal bloom in July 2008," *J. Mar. Syst.*, vol. 80, no. 1/2, pp. 135–142, Feb. 2010, doi: 10.1016/j.jmarsys.2009.02.010.

[13] S. Wang and D. Tang, "Preliminary remote sensing observation of sea surface temperature increase during *Ulva prolifera* blooms," *Aquatic Ecosyst. Health Manage.*, vol. 17, no. 3, pp. 299–304, Jul. 2014, doi: 10.1080/14634988.2014.944471.

[14] Y. Xiao, J. Zhang, and T. Cui, "High-precision extraction of nearshore green tides using satellite remote sensing data of the Yellow Sea, China," *Int. J. Remote Sens.*, vol. 38, no. 6, pp. 1626–1641, Mar. 2017, doi: 10.1080/01431161.2017.1286056.

[15] W. Shang, Z. Gao, X. Jiang, C. Liu, and W. Gao, "Green tide disaster monitoring system based on multi-source data," *Proc. SPIE*, vol. 9975, Sep. 2016, Art. no. 99750O, doi: 10.1117/12.2235897.

[16] F. Xu *et al.*, "A UAV and S2A data-based estimation of the initial biomass of green algae in the south Yellow Sea," *Mar. Pollut. Bull.*, vol. 128, pp. 408–414, Mar. 2018, doi: 10.1016/j.marpolbul.2018.01.061.

[17] Q. Xing *et al.*, "Remote sensing of early-stage green tide in the Yellow Sea for floating-macroalgae collecting campaign," *Mar. Pollut. Bull.*, vol. 133, pp. 150–156, Aug. 2018, doi: 10.1016/j.marpolbul.2018.05.035.

[18] S. Cao, D. Hu, Z. Hu, W. Zhao, Y. Mo, and K. Qiao, "An integrated soft and hard classification approach for evaluating urban expansion from multisource remote sensing data: A case study of the Beijing–Tianjin–Tangshan metropolitan region, China," *Int. J. Remote Sens.*, vol. 39, no. 11, pp. 3556–3579, Jun. 2018, doi: 10.1080/01431161.2018.1444291.

[19] D. Blondeau-Patissier, J. F. R. Gower, A. G. Dekker, S. R. Phinn, and V. E. Brando, "A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans," *Prog. Oceanogr.*, vol. 123, pp. 123–144, Apr. 2014, doi: 10.1016/j.pocean.2013.12.008.

[20] Y. Wang, S. Ruan, T. Wang, and M. Qiao, "Rapid estimation of an earthquake impact area using a spatial logistic growth model based on social media data," *Int. J. Digit. Earth*, vol. 12, no. 11, pp. 1265–1284, Nov. 2019, doi: 10.1080/17538947.2018.1497100.

[21] W. Zhai and C. Cheng, "A long short-term memory approach to predicting air quality based on social media data," *Atmos. Environ.*, vol. 237, Sep. 2020, Art. no. 117411, doi: 10.1016/j.atmosenv.2020.117411.

[22] C. Zhang, C. Fan, W. Yao, X. Hu, and A. Mostafavi, "Social media for intelligent public information and warning in disasters: An interdisciplinary review," *Int. J. Inf. Manage.*, vol. 49, pp. 190–207, Dec. 2019, doi: 10.1016/j.ijinfomgt.2019.04.004.

[23] B. Poblete, J. Guzman, J. Maldonado, and F. Tobar, "Robust detection of extreme events using Twitter: Worldwide earthquake monitoring," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2551–2561, Oct. 2018.

[24] P. Earle, M. Guy, R. Buckmaster, C. Ostrum, S. Horvath, and A. Vaughan, "OMG earthquake! Can Twitter improve earthquake response?" *Seismol. Res. Lett.*, vol. 81, no. 2, pp. 246–251, Mar. 2010, doi: 10.1785/gssrl.81.2.246.

[25] M. Imran, S. Elbassuoni, C. Castillo, F. Diaz, and P. Meier, "Extracting information nuggets from disaster-related messages in social media," in *Proc. 10th Int. Conf. Inf. Syst. Crisis Response Manage.*, Baden-Baden, Germany, May 2013. [Online]. Available: http://idl.iscram.org/files/imran/2013/613_Imran_etal2013.pdf

[26] J. H. Lee, I.-C. Pang, I.-J. Moon, and J.-H. Ryu, "On physical factors that controlled the massive green tide occurrence along the southern coast of the Shandong Peninsula in 2008: A numerical study using a particle-tracking experiment," *J. Geophys. Res.*, vol. 116, no. C12, Dec. 2011, Art. no. C12036, doi: 10.1029/2011JC007512.

[27] F. Bonjean and G. S. E. Lagerloef, "Diagnostic model and analysis of the surface currents in the tropical pacific ocean," *J. Phys. Oceanogr.*, vol. 32, no. 10, pp. 2938–2954, 2001.

[28] M. Bao, W. Guan, Y. Yang, Z. Cao, and Q. Chen, "Drifting trajectories of green algae in the western Yellow Sea during the spring and summer of 2012," *Estuarine, Coastal Shelf Sci.*, vol. 163, pp. 9–16, Sep. 2015, doi: 10.1016/j.ecss.2015.02.009.

[29] D. Li, Z. Gao, D. Song, W. Shang, and X. Jiang, "Characteristics and influence of green tide drift and dissipation in Shandong Rongcheng coastal water based on remote sensing," *Estuarine, Coastal Shelf Sci.*, vol. 227, Oct. 2019, Art. no. 106335, doi: 10.1016/j.ecss.2019.106335.

[30] X. Liu, Y. Li, Z. Wang, Q. Zhang, and X. Cai, "Cruise observation of Ulva prolifera bloom in the southern Yellow Sea, China," *Estuarine, Coastal Shelf Sci.*, vol. 163, pp. 17–22, Sep. 2015, doi: 10.1016/j.ecss.2014.09.014.

[31] S. Wang *et al.*, "Accurate Ulva prolifera regions extraction of UAV images with superpixel and CNNs for ocean environment monitoring," *Neurocomputing*, vol. 348, pp. 158–168, Jul. 2019, doi: 10.1016/j.neucom.2018.06.088.

[32] L. M. Scott and M. V. Janikas, "Spatial statistics in ArcGIS," in *Handbook of Applied Spatial Analysis*, M. M. Fischer and A. Getis, Eds. Berlin, Germany: Springer, 2010, pp. 27–41.

[33] T. Iwata, A. Shah, and Z. Ghahramani, "Discovering latent influence in online social activities via shared cascade Poisson processes," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Chicago, IL, USA, 2013, p. 266, doi: 10.1145/2487575.2487624.

[34] O. Komori, S. Eguchi, Y. Saigusa, B. Kusumoto, and Y. Kubota, "Sampling bias correction in species distribution models by quasi-linear Poisson point process," *Ecol. Inform.*, vol. 55, Jan. 2020, Art. no. 101015, doi: 10.1016/j.ecoinf.2019.101015.
[35] T. Nelson, "Statistical analysis of spatial and spatio-temporal point patterns, third edition," *Int. J. Geographical Inf. Sci.*, vol. 29, no. 4, pp. 690–691, Apr. 2015, doi: 10.1080/13658816.2014.937718.
[36] J. A. González, F. J. Rodríguez-Cortés, O. Cronie, and J. Mateu, "Spatio-temporal point process statistics: A review," *Spatial Statist.*, vol. 18, pp. 505–544, Nov. 2016, doi: 10.1016/j.spasta.2016.10.002.
[37] S. Zhang, X. Liu, J. Tang, S. Cheng, and Y. Wang, "Urban spatial structure and travel patterns: Analysis of workday and holiday travel using inhomogeneous Poisson point process models," *Comput., Environ. Urban Syst.*, vol. 73, pp. 68–84, Jan. 2019, doi: 10.1016/j.compenvurbsys.2018.08.005.
[38] P. J. Diggle, *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, 3rd ed. Boca Raton, FL, USA: Chapman & Hall/CRC, 2013.
[39] A. L. Thurman and J. Zhu, "Variable selection for spatial Poisson point processes via a regularization method," *Statist. Methodol.*, vol. 17, pp. 113–125, Mar. 2014, doi: 10.1016/j.stamet.2013.08.001.
[40] A. C. Linke *et al.*, "Dynamic time warping outperforms Pearson correlation in detecting atypical functional connectivity in autism spectrum disorders," *NeuroImage*, vol. 223, Dec. 2020, Art. no. 117383, doi: 10.1016/j.neuroimage.2020.117383.
[41] Y. H. Chan, "Biostatistics 104: Correlational analysis," *Singap. Med. J*, vol. 44, no. 12, pp. 614–619, 2003.
[42] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Evanston, IL, USA: Routledge, 1988.
[43] H. Akoglu, "User's guide to correlation coefficients," *Turkish J. Emergency Med.*, vol. 18, no. 3, pp. 91–93, Sep. 2018, doi: 10.1016/j.tjem.2018.08.001.
[44] M. Molahasani Majdabadi, S. B. Shokouhi, and S.-B. Ko, "Efficient hybrid CMOS/Memristor implementation of bidirectional associative memory using passive weight array," *Microelectron. J.*, vol. 98, Apr. 2020, Art. no. 104725, doi: 10.1016/j.mejo.2020.104725.
[45] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.
[46] R. Kumar *et al.*, "Revealing the benefits of entropy weights method for multi-objective optimization in machining operations: A critical review," *J. Mater. Res. Tech.*, vol. 10, pp. 1471–1492, Jan. 2021, doi: 10.1016/j.jmrt.2020.12.114.
[47] H. Song, Q. Liu, G. Wang, R. Hang, and B. Huang, "Spatiotemporal satellite image fusion using deep convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 821–829, Mar. 2018.
[48] C. Zhang, X. Xue, Y. Zhao, X. Zhang, and T. Li, "An improved association rule mining-based method for revealing operational problems of building heating, ventilation and air conditioning (HVAC) systems," *Appl. Energy*, vol. 253, Nov. 2019, Art. no. 113492, doi: 10.1016/j.apenergy.2019.113492.
[49] E. K. Buxton, S. Vohra, Y. Guo, A. Fogleman, and R. Patel, "Pediatric population health analysis of southern and central Illinois region: A cross sectional retrospective study using association rule mining and multiple logistic regression," *Comput. Methods Programs Biomed.*, vol. 178, pp. 145–153, Sep. 2019, doi: 10.1016/j.cmpb.2019.06.020.



Zhongyuan Wang is currently working toward the Ph.D. degree in cartography and geography information with the State Key Laboratory for Information Engineering in Surveying, Cartography and Geographical Information Engineering, Wuhan University, Wuhan, China.

His research interests include data fusion and algorithm study.



Zhixiang Fang received the M.Sc. degree in geodesy and surveying engineering and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2005 respectively.

He is currently a Professor with the State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing. His research interests include spatiotemporal modeling of human behavior, space-time GIS for transport, and intelligent navigation.



Yu Zhang received the Ph.D. degree in cartography and geography information from Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China.

His research interests include geographical information retrieval and geospatial data mining.



Zhanlong Song received the bachelor's degree in computer science and technology from the University for science and Technology, Beijing, China in 2011.

He is currently an Engineer with Beijing Global Safety Technology Co. Ltd., Beijing, China. His research interests include web application development and software development management.