Journal of Environmental Management 203 (2017) 87–97

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

Monitoring the environment and human sentiment on the Great Barrier Reef: Assessing the potential of collective sensing

Susanne Becken [a,*], Bela Stantic [b], Jinyan Chen [b], Ali Reza Alaei [a], Rod M. Connolly [c]

[a] Griffith Institute for Tourism, Griffith University, Gold Coast 4222, Australia
[b] Institute for Integrated and Intelligent Systems, Griffith Sciences, Griffith University, Gold Coast 4222, Australia
[c] Australian Rivers Institute – Coast and Estuaries, School of Environment, Griffith University, Gold Coast 4222, Australia

ARTICLE INFO

Article history:
Received 28 February 2017
Received in revised form
28 June 2017
Accepted 3 July 2017

Keywords:
Environmental monitoring
Social media
Twitter
Great Barrier Reef
Sentiment analysis

ABSTRACT

With the growth of smartphone usage the number of social media posts has significantly increased and represents potentially valuable information for management, including of natural resources and the environment. Already, evidence of using 'human sensor' in crises management suggests that collective knowledge could be used to complement traditional monitoring. This research uses Twitter data posted from the Great Barrier Reef region, Australia, to assess whether the extent and type of data could be used to Great Barrier Reef organisations as part of their monitoring program. The analysis reveals that large amounts of tweets, covering the geographic area of interest, are available and that the pool of information providers is greatly enhanced by the large number of tourists to this region. A keyword and sentiment analysis demonstrates the usefulness of the Twitter data, but also highlights that the actual number of Reef-related tweets is comparatively small and lacks specificity. Suggestions for further steps towards the development of an integrative data platform that incorporates social media are provided.

1. Introduction

We are living in a networked society, and the use of mobile Internet is a recent phenomenon that has experienced exponential growth. With the growth in Internet subscriptions and smartphone usage, the engagement with social media has increased as well. Smartphones themselves are tracking devices, and the information shared through social media – especially when it is spatially and temporally tagged – bears great potential for monitoring environmental changes (Shook and Turner, 2016). The possibility of using social media posts as a tool to access diverse and unique information provided by millions of 'connected' citizens or 'human sensors' for environmental management purposes will be introduced in this work.

Tapping into social media can greatly enhance existing approaches where environmental data are more purposefully collected by citizens, for example in the areas of biodiversity and conservation (see Couvet et al., 2008). One of the first applications of using social media sensors has been in disaster management. The analysis of 10 million Twitter posts in the aftermath of Hurricane Sandy in New York in 2012 demonstrated that tweets reported damage faster and more accurately than the National Federal Emergency Management Agency (Bohannon, 2016). Capitalising on the real-time spread of information via social media, the U.S. Geological Service has now complemented its network of seismological sensors with data mining of Twitter feeds (Meyer, 2015). The development of social media in enhanced decision making systems is advancing rapidly, both in response to environmental shocks and longer term pressures (Shook and Turner, 2016).

Applications in environmental management and ecological changes are less established, but well-known natural attractions that are visited by large numbers of people could be well suited. The Great Barrier Reef (GBR) in Australia is such an attraction. It is one of the world's most iconic World Heritage Areas, and is a biodiversity hotspot and showpiece of the Australian tourism industry. Over 2.2 million international and 1.7 million domestic visitors travel to the GBR every year (Tourism Research Australia, 2015). Additionally, there are more than one million people living in the region, the majority of whom are active users of the Reef and its adjacent beaches (Deloitte Access Economics, 2013). The GBR therefore represents an excellent opportunity to investigate

* Corresponding author.
E-mail addresses: s.becken@griffith.edu.au (S. Becken), b.stantic@griffith.edu.au (B. Stantic), jinyan.chen@griffith.edu.au (J. Chen), a.alaei@griffith.edu.au (A.R. Alaei), r.connolly@griffith.edu.au (R.M. Connolly).

http://dx.doi.org/10.1016/j.jenvman.2017.07.007
0301-4797/

whether social media users provide valuable information on the environment and related experiences. Sentiment relates to valence and reflects underlying emotions, broadly classified into positive, neutral and negative. Sentiment can be extracted from social media statements through the use of computational linguistics and natural language processing.

Environmental monitoring has become increasingly critical, as the GBR ecosystem has undergone significant change and decline in its ecological quality (Great Barrier Reef Water Science Taskforce & Department of Environment and Heritage Protection, 2016). The 2016 coral bleaching event resulted in a mortality of 29% of shallow water corals (Great Barrier Reef Marine Park Authority [GBRMPA], 2017). Bleaching is continuing in 2017 due to warm water temperatures and GBRMPA staff are manually screening Instagram photos to obtain a first sense of the spatial extent of bleaching (personal communication Chris Jones, GBRMPA, May 2017). In the face of multiple interacting and cumulative stress factors that compromise Reef health, the GBRMPA is now working towards an integrated monitoring program to help evaluate progress towards long-term sustainability targets. The goal is to better integrate the many existing monitoring programs and also address gaps by formally implementing non-traditional approaches such as citizen science (Addison et al., 2015).

Using the GBR as a prominent case study, this research addresses several research questions. First, is the scale of tweets and their spatial distribution sufficient for the collection of relevant information on the GBR environment? Second, who are the providers of information, and third, is the information that can be extracted from tweets useful for environmental monitoring? Since a ‘human sensing’ approach to environmental monitoring is new, the findings from this research and the discussion focus on drawing out valuable insights and ‘lessons learned’ about the benefits and challenges associated with using social media data for environmental purposes.

2. Human sensing and monitoring

2.1. Applications of human sensing

Managers of environmental resources are beginning to consider the use of new types of data that originate from the exponential uptake in mobile smartphone technology, people's willingness to share content, and the ability to track users via their Geographic Positioning System (GPS); rapid progress in data science has generated considerable interest in using this new information source for a range of purposes (Connors et al., 2012; Shook and Turner, 2016). Research in this area is building on earlier developments in the area of ‘market sensing’ or ‘social listening’ (Rao, 2014). In particular, the tourism industry has advanced both theory and practice on how to use online generated content to refine product development, marketing and customer experience (Leung et al., 2013). It is now recognized that such non-traditional approaches present inexpensive means for gathering rich, authentic, and unsolicited data on people's perceptions and experiences (O'Leary, 2011).

There are several examples of how those involved in environmental management have accessed large numbers of individual observations on specific phenomena. For example, Kirilenko et al. (2015) analysed the climate change discourse evident from over 2 million tweets in 157 cities in the United States. It was found that both deviations from ‘normal temperatures’ and climate change coverage in the mass media had a significant influence on the number and content of climate change-related tweets. Also in the United States, researchers analysed photo imagery uploaded on Flickr, a photo-sharing website, to replace costly visitor surveys for monitoring recreational visitation to lakes. The photos were used to generate the metric of ‘photo-user-days’, which was then employed in the development of a visitation model that links water clarity with visitation levels (Keeler et al., 2015). The research provided robust evidence that social media data can be used in human-environment research. Future research could focus on the photo content to examine ecological changes. Building on pioneering work by Keeler et al. (2015), scientists from The Nature Conservancy used Flickr photos to derive visitation numbers to coral reefs (The Nature Conservancy, 2017). The resulting global estimates of the economic value of reefs, and the generation of an interactive map through Mapping Ocean Wealth, illustrate the benefits of using social media data for the purpose of managing natural resources.

Collecting time-stamped and spatially relevant data from social media has been most advanced in the area of disaster management (Crooks, Croitoru, Stefanidis and Radzikowski, 2013; Steiger et al., 2015). Researchers and emergency management organisations found that tapping into the subjective information provided by citizens can greatly enhance rapid response and decision making during acute crises (Chae et al., 2014). One application is to assess the extent of damage and evacuation in near-real time (Bohannon, 2016; Crooks et al., 2013; Schnebele and Cervone, 2013). Social media has also been identified as an effective channel to communicate relevant information to affected communities; in particular through users with a large followership. Researchers have also explored the potential of building integrated disaster management systems that combine informal social networking and formal disaster communication technologies into one centralised platform (Avvanuti et al., 2016). As a result, a two-way communication channel between emergency services and affected people within or even outside the area can be established.

The use of ‘collective knowledge’ (Vivacqua and Borges, 2012) in emergency situations is particularly effective because people tend to engage with their online social networks in extreme situations. Researched examples include Twitter messages posted during and after the Boston marathon explosions (Cassa et al., 2013), terrorism attacks in Jakarta and Mumbai (Cheong and Lee, 2011), and a terror act in a shopping mall in Kenya (Simon et al., 2014). Similarly, data from social media have been found useful to understand outbreaks and spreads of infectious diseases (Brownstein et al., 2008).

Whilst extreme situations lend themselves for the exploitation of social media data, the opportunities for longer term monitoring of the environment have been investigated to a lesser extent. One exception is urban air quality, possibly because changes in quality are noticeable and of public concern, and cities provide a critical mass of social media users who can provide sufficient volumes of ‘measurements’. Riga and Karatzas (2014), for example, analysed tweets and developed a Self-Organizing Map that tracks the environmental loads and air quality affecting people's lives. Another recent study on the use of Twitter data for conservation purposes is noteworthy. Daume (2016) analysed 2842 tweets that made references to particular invasive alien species. The findings confirmed that Twitter can be a useful source of information on species occurrence, but also on human perceptions.

Other approaches to utilising ‘human sensors’ use a more structured approach. For example, people are encouraged to engage in a process of voluntary provision of information on specifically designed web-based platforms or citizen-based data collection initiatives in the field. Comparisons of data collected by citizens, for example related to wildlife observations along roads, with those from scientists have demonstrated good overlap (Paul et al., 2014). A recent analysis of amateur weather station networks highlights the potential of citizen-collected data and the opportunity to link these with traditional decision maker networks

(Gharesifard et al., 2017).

2.2. Monitoring in the Great Barrier Reef region

The Reef 2050 Long-Term Sustainability Plan by the Australian and Queensland Government (2015) was prepared in response to the UNESCO World Heritage Committee's recommendation to ensure protection of the Reef's Outstanding Universal Value. The related Implementation Strategy identifies governance arrangements, as well as the need for an Integrated Monitoring and Reporting Program. The Reef 2050 Integrated Monitoring and Reporting Program Strategy (Commonwealth of Australia, 2015) foresees to create better alignment between the current 90 or more monitoring programs that are operating in the GBR Marine Park. The existing monitoring programs cover a wide range of areas, including the marine environment, water quality, tourism and recreation, fisheries, and socio-economic trends such as community benefits. The strategy specifically mentions citizen science monitoring, referring to programs such as CoralWatch or Seagrass-Watch.

GBRMPA, the managing authority, has engaged in various forms of human sensing for some time. The Eye on the Reef program enables visitors and operators to contribute information about reef health, marine animals and incidents. The program facilitates the contribution of data through various platforms. At the least formal level, visitors to the Reef can provide information through a mobile App or online system. The App is used to report sightings of particular species and upload photos. The information collected contains the particular content of interest (e.g. a particular species), the time and the location it relates to. This Volunteered Geographic Information (Connors et al., 2012; Resch, 2013) generates new people-moderated data that can be added to professional monitoring programs. In addition, Reef tourism operators are contributing through the Rapid Monitoring Survey. This part of the Eye on the Reef program requires more experience and training, as an underwater monitoring slate is completed and submitted to the database through an online portal. The highest level of data accuracy is achieved through the Tourism Operators Weekly Monitoring Survey, which demands ongoing commitment to monitoring environmental indicators in the same location. In those two latter approaches the providers of information are more than coincidental sensors of the environment at a given time (Resch, 2013) and the quality of the information is likely to be superior.

At this point, the GBR has no mechanism to capture broader ‘collective knowledge’ on the Reef. It is the aim of this research to provide a baseline assessment of the suitability of Twitter feeds for the purpose of developing social media platform that enhances traditional monitoring of the GBR. If successful, a framework or architecture similar to the one proposed by Avvanuti et al. (2016) in the context of emergency management could be developed.

3. Method

Twitter was used as the source of data, because it is a relatively commonly used platform that makes up about 4% of the global social media activity (Chaffey, 2016). Facebook's market share is higher (18%), but the content is not publicly available. For Twitter, a sample of at least 1% of the equivalent of tweets posted daily (out of 500,000,000 per day, Twitter, 2016) is freely available for analysts (Avvanuti et al., 2016; Crooks et al., 2013). The length limit of 140 characters per tweet means that processing is simplified in some ways, but analysis might be challenging because of limited information contained in the short text. Tweets are accompanied with background information on the user, which is useful for interpreting the content. As in earlier studies (Steiger et al., 2015), this research entails both a spatial and semantic analysis of tweets. This section details how the data were accessed and stored, and what procedures were employed to analyse user statistics and content.

3.1. Accessing Twitter data

To retrieve data, we employed an online streaming approach. Specifically we used a public Twitter API with restrictions to capture geo-tagged tweets posted from the GBR region. Geo-tagged tweets are a subsample of tweets associated with explicit geographic coordinates measured by either an exact coordinate or an approximate coordinate (polygon). For a tweet associated with an exact location, the coordinates are obtained either based on GPS embedded in mobile devices, or on the IP location of the computer located to the nearest address (Hawelka et al., 2014). In the case of a tweet associated with a polygon, the polygon is created based on either the place (place_id) that the sender explicitly specified when the tweet was posted, or on the default place (place_id) chosen by Twitter from the user profile location. The exact way in which Twitter assigns a polygon is not fully transparent, which poses a limitation to this research (Steiger et al., 2015).

Two types of errors emerge. First, tweets from people who visited the GBR region and have chosen a high level of privacy (“Add a location to my Tweets” is not selected or “geo-enabled = FALSE”) will not be recognized by our data collection approach. This error is a Type II error as a number of tweets remain undetected, and information that could have been useful cannot be incorporated. The main implication is that the data volume is smaller, but there is no reason to assume systematic biases. The second error relates to tweets that Twitter believed to originate from the GBR region, when they were actually posted from another region. This might occur when an account holder is registered in the GBR bounding box, but travels outside the region and has all location enabling services turned off. Twitter is likely to assign every tweet to the person's location of account. This Type I error is slightly more problematic because it adds irrelevant tweets to the sample However, subsequent filtering (e.g. for keywords) is likely to reduce the impact of this error.

To determine an approximate region of the GBR for data collection a rectangular bounding box was considered (Southwest coordinates: 141.459961, −25.582085 and Northeast coordinates: 153.544922, −10.69867). The bounding box does not perfectly overlay what is normally considered as the ‘Great Barrier Reef region’, either geographically or administratively. However, most data come from the coastal areas of the GBR region, with only a few ‘touching’ the boundary. This resulted in the download of about 1500 tweets per day, although as a result of server failure, several days are missing. For the purpose of this present research these missing days do not present a problem, however, future real-time assessments will require stable systems (including back up for power outage) to ensure no loss of data. Data are stored in a NoSQL MongoDB database, which is located on a cluster computer with a Hadoop architecture. Each tweet in the database contains Metadata (Steiger et al., 2015), including the content of the tweet, language, location where account was opened, and place from which the tweet was sent (Table 1).

3.2. Keyword analysis

Twitter users send tweets for a wide range of reasons, and it should not be surprising that only a small number of tweets refer to the GBR. It is therefore of critical importance to filter the large number of tweets to extract messages of interest. Thus, a framework of categories and key terms was developed to filter those tweets that might provide insight into the marine environment. Five keyword categories were used (for a full list of keywords, see

Table 1
Relevant data field stored in this project's Twitter database.

Variable Name	Variable Label
username	User name
Id	Respondent ID
userstatuses_count	Count of User Statuses
text	Tweet text
Lang	Language of Tweet
timestamp_ms	Time stamp of tweet
created_at	Time tweet created (e.g. Tue Mar 29 22:57:46 + 0000 2016)
Placename	Place tweet created (short)
placefull_name	Place tweet created (full name, location hierarchy)
usertime_zone	Users time zone setting (Account details)
Userlocation	Users specified location (Account details)
usercreated_at	When user created Twitter account (Account details)
userfollowers_count	Count of Users Followers (Account details)
Userlang	Users specified language (Account details)

Table 3):

- Locations (e.g. Cairns, Townsville)
- Activities (e.g. swim, snorkel, dive)
- Marine life (e.g. fish, turtle, shark)
- Water (e.g. clarity, visibility)
- Coral (e.g. bleach, white, colourful)

All keywords were extracted using a case insensitive search technique, and variations of the same word (e.g. 'dive', 'diving') were compiled as the same keyword. Numbers of occurrences for each keyword were counted.

3.3. Sentiment analysis

In addition to keyword frequencies, tweets were analysed with regards to their positive or negative polarity. Sentiment analysis is regarded as an efficient method for analysing social media content. Scoring sentiment is an analytical approach that converts subjective and unconstructed text into constructed data. The purpose is to extract information that reveals critical events and assists in determining the emotional tone behind textual data in order to gain an understanding of opinions. Sentiment analysis is a challenging task because of the need to simplify complex content, deduct a polarity from short sentences, interpret emoticons, and capture meaning despite grammatical and syntactical mistakes. The analysis is further complicated by a tendency to use abbreviated language conventions, by the use of slang or other linguistic tools such as irony or sarcasm (for a review of sentiment analysis approaches, see Alaei et al., 2017; Hutto and Gilbert, 2014).

In this work we accepted a recently proposed approach for sentiment analysis (Ribeiro et al., 2016) that was specifically developed for the analysis of social media text. Valence Aware Dictionary for Sentiment Reasoning (VADER) is a rule-based model that combines a general lexicon and a series of intensifiers, punctuation transformation, emoticons, and many other heuristics to compute sentiment polarity of a review or text. The VADER sentiment lexicon is composed of more than 7000 items along with their associated sentiment intensity measures, validated by humans. The sentiment score ranges from minus one (negative) to plus one (positive), with the middle point being considered 'neutral'. The VADER only provides sentiment for English tweets, and for text written in other languages it assigns neutral polarity. Tweets were therefore kept in the database as they still provided insight into volumes and account holders, but were not part of the sentiment analysis. There is potential in future to make use of translational application program interfaces (API) to assess non-English tweets.

4. Results

Over the period from the 18th March to 31st October 2016, a total of 208,525 tweets that were identified by Twitter as posted from the GBR region were stored in a database. Relevant Metadata were examined to help answer the question on where tweets were posted from and who was providing the 'collective knowledge' of the Reef. This is followed by a keyword and sentiment analysis to assess relevance of the information provided for an environmental monitoring system.

4.1. Where tweets are posted from

The majority of tweets were posted from inexact locations within the GBR bounding box. Only 16.7% of the tweets contain coordinate points, which represent exact locations (longitude and latitude). These georeferenced tweets were mainly posted from the coastal zones, but also from areas inland and on the Reef (possibly from boats or islands; Figs. 1 and 2). The heat maps show the number of tweets posted regardless of their content (i.e. relevant to the Reef or not) and reflect the main population centres and tourist destinations of the Cairns region in the North (Wet Tropics: 237,351 population), Townsville (Burdekin: 222,116), Airlie Beach (MacKay and Whitsundays: 131,537), and Rockhampton in the more Southern parts of the GBR (Fitzroy: 227,830). The finer detail of tweet locations in the Cairns region shows islands visited by tourists on snorkel or diving trips. It is possible that tweets posted from these locations contain valuable information about the GBR environment.

4.2. Who posts tweets?

Of the total of 208,525 tweets in our database, 150,625 tweets (72%) were from user accounts with meaningful locations provided in their account profiles. Considering only those tweets, there were 1236 unique users who were registered in the GBR region (Table 2). This was the smallest group of account holders, with the majority coming from other specified locations in Australia. There are some uncertainties, as for example 13.7% of Twitter users registered their account 'Australia', and another 5.5% only specified 'Queensland' as their location; thus not providing a clearly identifiable location. Over one fourth of the unique account holders were from overseas.

When looking at the number of tweets posted (rather than unique users), a total of 64.2% of tweets came from visitors (Table 2), highlighting the potential value of social media analysis in tourist destinations that are otherwise relatively sparsely populated. By volume, most tweets came from people registered in

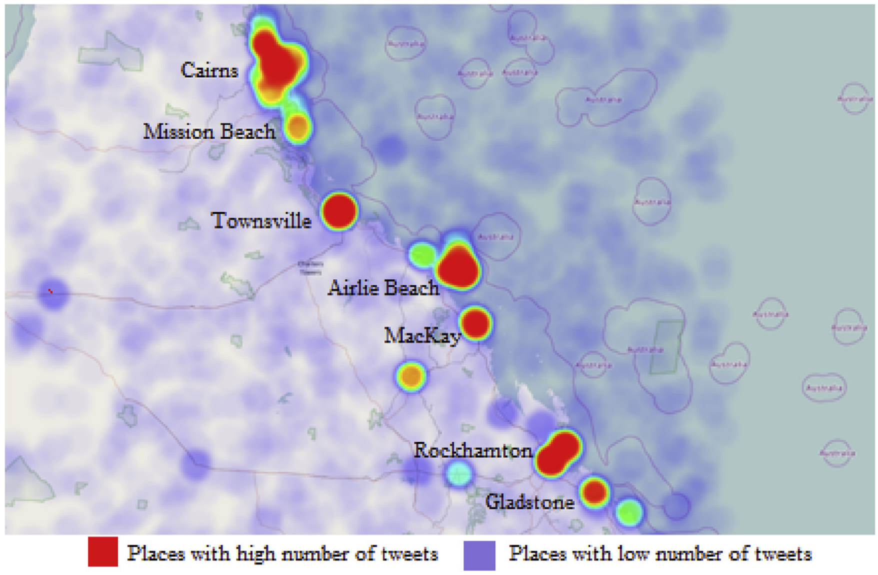

Fig. 1. Heat map of geo-referenced tweets showing where tweets were posted from between March and October 2016.

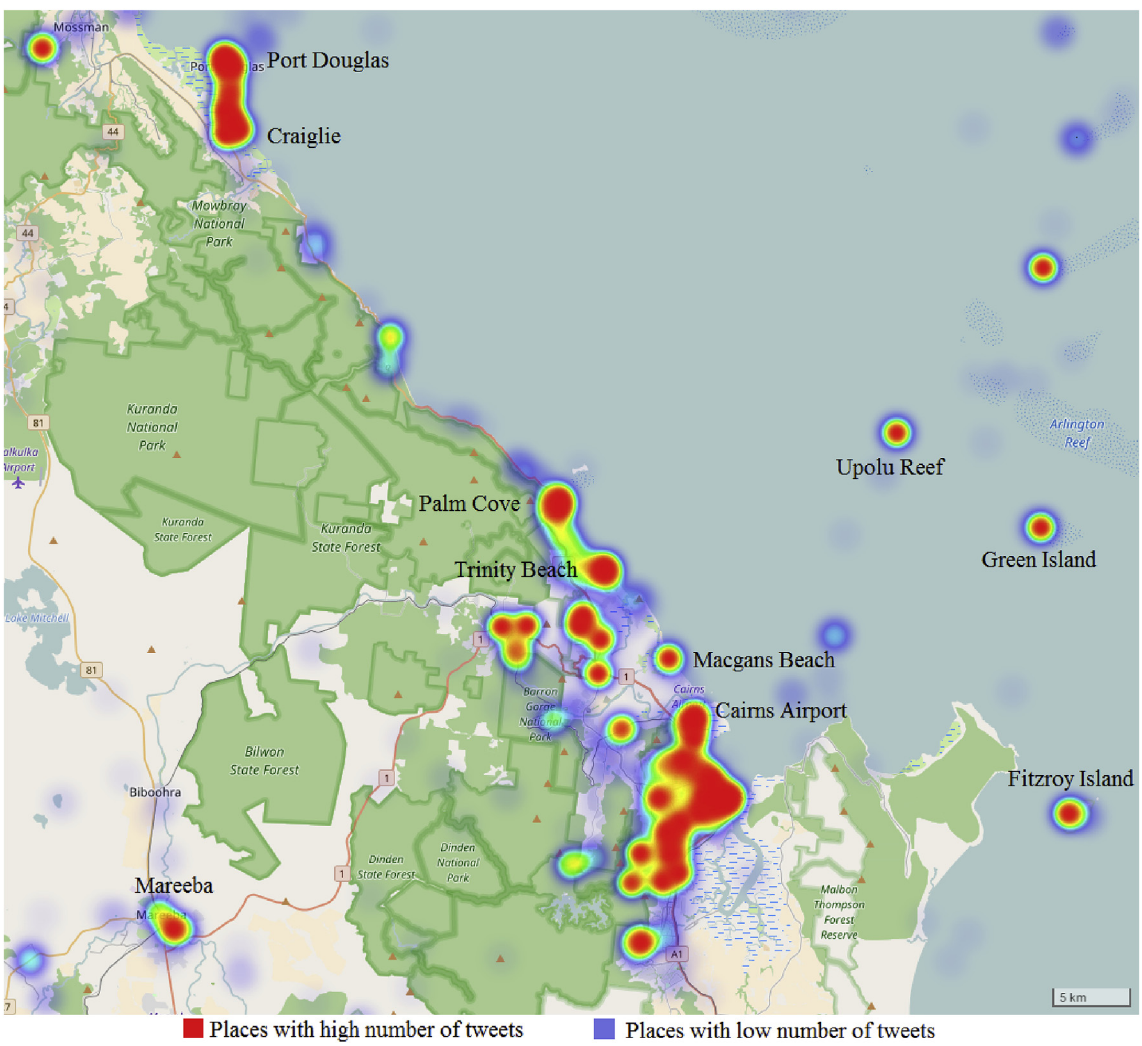

Fig. 2. Detailed heat map of geo-referenced tweets in the Cairns region posted between March and October 2016.

Table 2
Statistics on the location of account holders, both in terms of unique users and volume of tweets.

Geographic region	Number	Proportion (%)
Unique registered users		
GBR	1236	11.9%
Australia (other than GBR)	6364	61.3%
International	2776	26.8%
Tweets sent by account holders from…		
GBR	53,955	35.8%
Australia (other than GBR)	75,697	50.3%
International	20973	13.9%

Table 3
Keyword numbers and examples of tweets that were identified based on keyword categories.

Reef locations mentioned in tweet	N	Reef activities	N	Specific species	N	Water	N	Coral attributes	N
Cairns	5052	Dive	545	Fish	804	Blue	45	Bleach/bleaching	41
Townsville	4421	Swim	518	Shark	525	Turquoise	3	Green	8
Rockhampton	747	Boat	400	Coral	307	Pristine	3	Algae	3
Hamilton Island	387	Snorkel	313	Dolphin	306	Dirty	1	White	3
Airlie Beach	321	Sail	282	Turtle	269	Clear	19	Damaged	2
Daintree	275	Marine	183	Whale	237	Mud(dy)	4	Dirty	2
Magnetic Island	275	Scuba	179	Nemo	140	Clean	4	Dead	1
Whitehaven Beach	225	Paddle	59	Dugong	31	*Beautiful/beauty*	*40*	Pristine	1
Whitsunday Islands	189			Clownfish	28	*Fresh/freshwater*	*40*	Brown	1
Mission Beach	85			Jellyfish	18	*Amazing*	*17*	*Climate change*	*18*
Green Island	84			Anemone	17	*Good*	*26*	*Scientists*	*8*
Cooktown	70			Stingray	15	*Best*	*15*	*Beautiful/beauty*	*8*
Fitzroy Island	57			Starfish	14	*Bad*	*6*	*Coal*	*7*
Daydream island	49			Trout	14	*Dead*	*3*		
Lady Elliot, Heron, Lady Musgrave Island	29			Wrasse, Grouper	12				
Example tweets									
Cheers to you Lady Elliot Island, you've been amazing.		*A very happy travel monkey - Snorkelling on the Great Barrier Reef @ Michaelmas Cay, Great.*		*Last night we all went in the ocean and fed the wild dolphins it was amazing although one nipped me was also a wobbegong shark (harmless)*		*A giant #turtle pops out of the water to say hi #greatbarrierreefâ€¦* https://t.co/uxSlDaCHXp		*Sorry Reef, when it's coal vs coral Barnaby will choose coal every time.*	
Airlie Beach is definitely one of the best East Coast stops so far.		*Turns out the Great Barrier Reef is a pretty good location for your first snorkelling Sesh.*		*OMG I'VE JUST SEEN A MUM AND A BABY DOLPHIN SWIM AROUND MY BOAT THIS IS THE BEST DAY*		*The average water temperature on Whitehaven Beach in the Whitsunday Islands is 26 Degrees. Fancy*		*Watching the stages of #coralbleaching go the wrong way. Stressed to fluorescent, bleached to algae.*	

Townsville (13.9%), Australia (13.9%), Cairns (12.2%), Queensland (10.5%), Sydney (5.3%), Melbourne (4.9%), Mackay (4.1%), and Brisbane (4.1%). Internationally, tweets from the UK were most common (3.4%), with others from the USA, Papua New Guinea, Canada, New Zealand, and Finland. In line with the majority of account holders, the main language of tweets was English (79.9%), followed by tweets in Japanese (5.1%), Spanish (1.4%), Tagalog (Philippines) (1.4%), French (1.1%), Indian (0.8%), Portuguese (0.8%), and Russian (0.6%). A total of 25 languages was detected.

4.3. Keywords analysis

The analysis of keywords revealed that only a small proportion of tweets related to content of potential relevance to the marine environment of the GBR. For example, only 0.6% of all tweets mentioned the keywords ‘water’ or ‘coral’. The location keywords were more commonly represented in the database of tweets. A proportion of 5.9% of all retrieved mentioned any of the keyword locations. Table 3 provides more details on the exact keywords and the associated number of tweets. It also presents examples of tweets for each keyword. For tweets containing the words water or coral, additional keywords were identified by counting frequent words associated with any tweet that either contained the word water or coral. These are listed in italics.

While people use Twitter to communicate their perceptions of particular places or activities (Table 3), it is also clear that the information is scant and often unspecific. Even tweets identified as relevant were not always useful in terms of environmental monitoring. For example, the tweet “*So amazing to be up close with some of the most beautiful fish in the world*” may be useful in understanding subjective experiences, but the content of the tweet reveals little about the environmental conditions. Rare examples of more informative tweets include the following one from Fitzroy Island, close to Cairns: “*Victory in the water today! Spotted frog fish, cuddle fish, sea turtle, string ray, and sharks!â€¦* https://t.co/fgGCcVTap9”. This tweet also contained a link to Instagram, which could provide further cues about the marine environment through photographic images.

4.4. Sentiment analysis

4.4.1. Overall sentiment

The VADER sentiment analysis technique classified over half of the tweets into either positive (37.9%) or negative (15.0%) sentiments, with 47.1% interpreted as neutral. The sentiment for tweets from the GBR bounding box fluctuated only slightly over time (see continuous lines in Fig. 2). Sentiment variations were more pronounced when analysing a specific location, for example the Cairns region (see dotted lines in Fig. 3). The middle lines shows the average score of positive and negative tweets, excluding those tweets that were classified as neutral. The average score tends to be positive and ranges between 0.254 in May and 0.275 in September. Those tweets mentioning Cairns, were more positive in the months of June and August than in April and July (Fig. 3).

4.4.2. Location sentiment

Sentiment associated with different locations in the GBR region

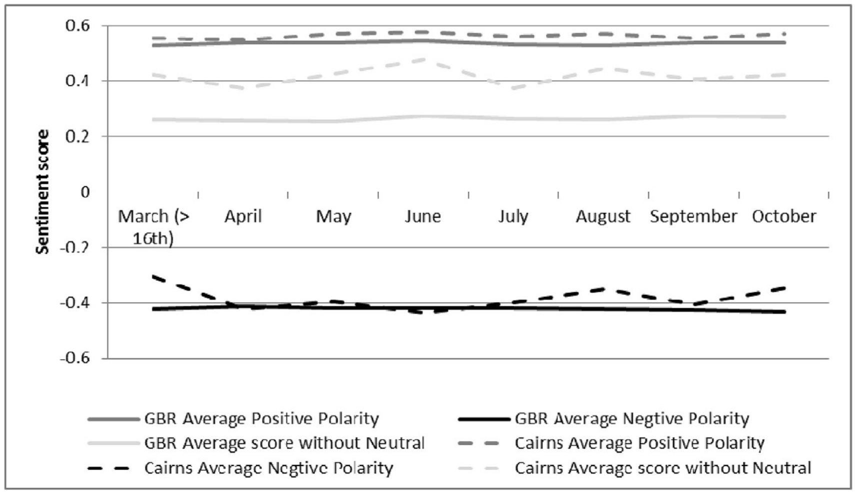

Fig. 3. Sentiment analysis of tweets for the GBR region in general and for tweets mentioning Cairns.

varies. Fig. 4 shows the number of tweets with positive, neutral or negative sentiment by location. In addition, Fig. 5 presents the average score of those tweets that were classified as either positive (with a maximum of 1) or negative (with a minimum of −1), and the average of both positive and negative tweets. Neutral tweets are not included in this visualisation. It is prudent to focus on those locations that have sufficiently large volumes of tweets. Cairns, for example, was mentioned in 5025 tweets and has a proportion of 38.9% of positive tweets and 7.2% of negative tweets. The average positive sentiment score is +0.57 (Fig. 5). In comparison, tweets about Green Island — an important tourist destination that was mentioned in 84 tweets — were more positive with 42.9% being classed as positive and an average positive sentiment score of +0.62. Lady Musgrave Island (N = 5 tweets, average score +0.62) and Heron Island (N = 16 tweets, average score +0.50), stand out, because they only attracted positive tweets; however, tweet numbers are too low to draw robust conclusions. The largest proportion of tweets classified as 'neutral' was observed for Cooktown (77.1%), Townsville (69.3), Rockhampton (66.0%) and Hamilton Island (63.3%). For the tourist destinations of Cooktown and Hamilton Island this could be a 'red flag' that visitors are not overly positive with the place, including possibly the marine environment.

4.4.3. Reef-related activities

Based on the assumption that people who engage in Reef-based activities might provide important clues on the environment, relevant tweets were extracted and analysed. A total of 2479 tweets that mentioned some kind of water-born activity were identified. Of these, only 10.4% were negative and 37.0% neutral. Thus, overall, the tweets reflect positive experiences with the Reef. Diving (represented by the keywords 'dive' with N = 545 and 'scuba' with N = 179) and 'snorkelling' (N = 313) are the key activities that might provide particular insight into the marine conditions. Tweets that mentioned 'snorkelling' and 'scuba' were largely positive (63.9% and 68.7%, respectively compared with 50.3% for 'diving') and had a high overall sentiment score of +0.55 and + 0.46. Diving only achieved an overall score of +0.36. Fig. 6 shows the positive, negative and overall scores for all identified key activities.

4.4.4. Marine life and water

Sentiment polarities of tweets containing keywords of marine life on the GBR show a small number of negative tweets (309, or

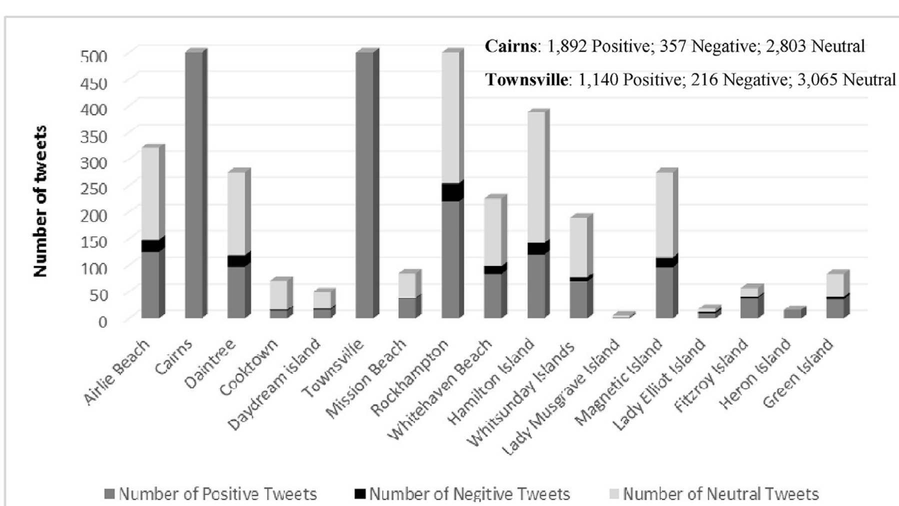

Fig. 4. Twitter volumes for key locations mentioned in tweets. Note that Cairns and Townsville are not fully displayed because of their large volume which would affect the scale on the y-axis.

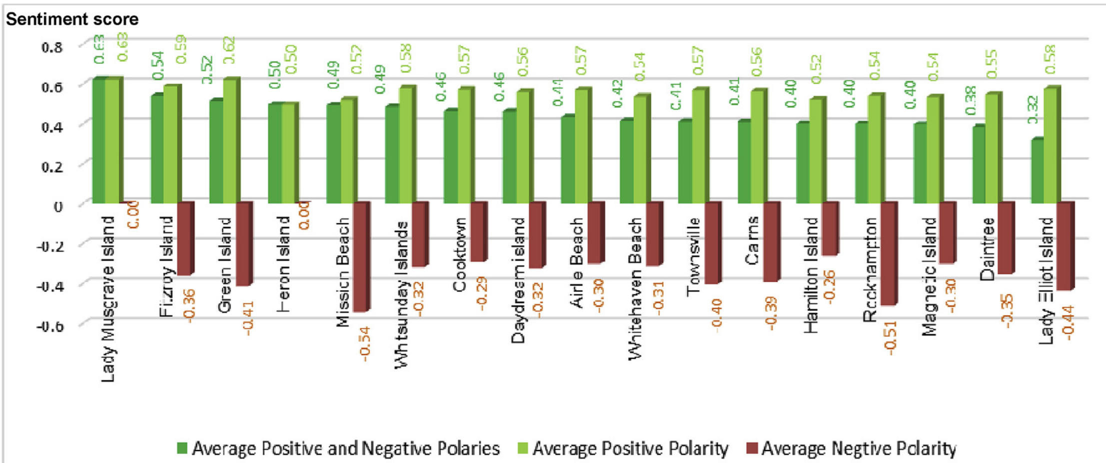

Fig. 5. Comparison of the sentiment polarities between key locations mentioned in the tweets. The results are ordered based on average positive and negative polarities.

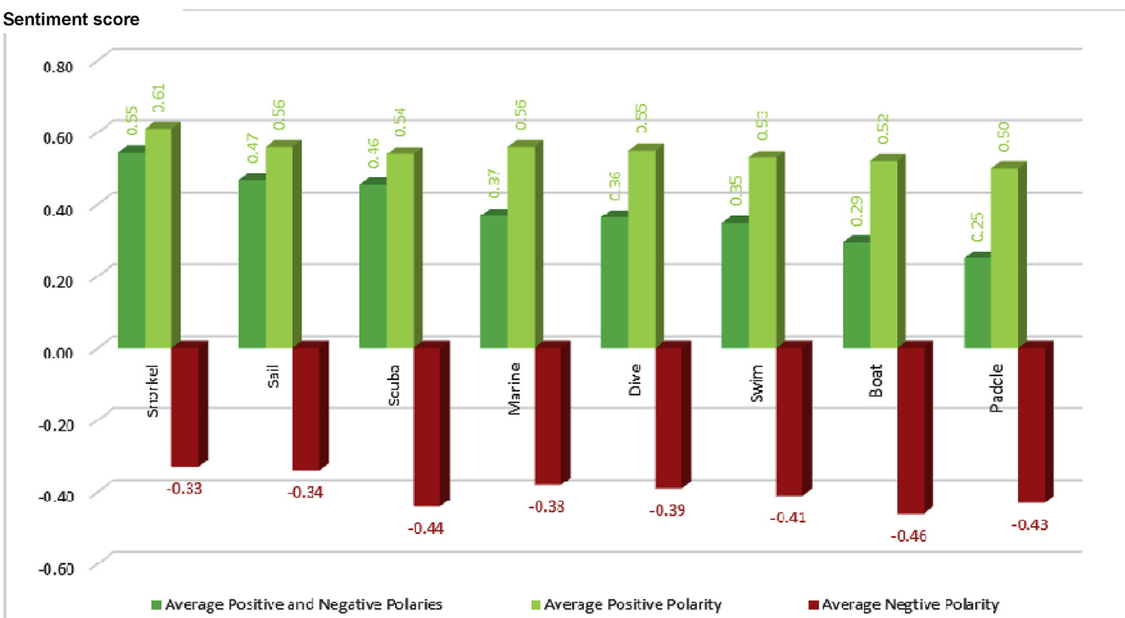

Fig. 6. Comparison of the sentiment polarities between activity keywords mentioned in the tweets. The results are ordered based on average positive and negative polarities.

11.3%, of all 2737 tweets related to marine life keywords (Fig. 7). A total of 1005 (or 36.7%) were neutral. The largest number of tweets related to 'fish' (N = 804), 'shark' (N = 525), 'coral' (N = 307) and dolphin (N = 306). Whilst there were more positive tweets (37.5%) than negative ones (13.0%), the ratio of positive to negative scores is slightly less favourable than for other keywords. The overall average score is therefore relatively low (0.32). Dolphins, by far, attracted the largest proportion of positive tweets (80.7%). Tweets that mentioned starfish (N = 14), trout (N = 14) and whale (N = 237) achieved the highest overall positive sentiment polarities. In contrast, Clownfish (N = 28), Jellyfish (N = 18) and Dugong (N = 31) attracted the most negative overall scores. Further examination is required to understand the underlying reasons.

In addition to marine species, tweets were filtered for keywords that related to 'water'. Surprisingly, the numbers of tweets were quite low (Table 3). For this reason, the sentiment analysis was only performed on words that contained the word 'water' rather than specific aspects. The word 'water' was mentioned in 1354 tweets, of which 39.2% were positive and 12.6% were negative. The average positive polarity was 0.527 and slightly lower than the polarities associated with keywords analysed further above. The average polarity across both negative and positive tweets was also comparatively low, but still positive at 0.30.

5. Discussion

This research gave an insight into the possibility of using Twitter data to enhance the GBRMPA's integrated monitoring system (Addison et al., 2015). A large number of tweets posted from within the region was captured and stored in a database. During the period of investigation between March and October 2016, about 1000–1500 tweets were saved per day. Tweets were mostly posted from the urban centres along the coast (e.g. Townsville, Cairns), but also from areas on the water (e.g. from boats) and on islands. The Twitter heat maps provide important context on where the information is coming from and highlight that, in response to the first

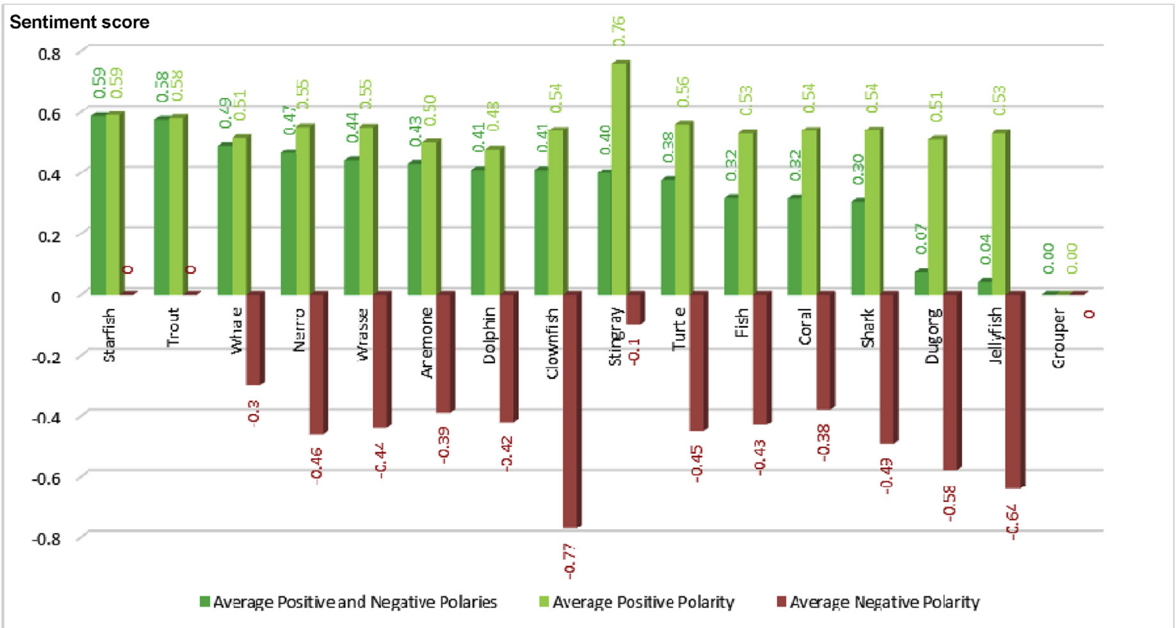

Fig. 7. Comparison of the sentiment polarities between marine life keywords mentioned in the tweets. The results are ordered based on average positive and negative polarities.

research question, the volume of tweets and their distribution gives confidence in using this data source for environmental management purposes. As evident in this present research, however, further research is necessary to improve the accuracy of location. At present only 17% of tweets are geo-referenced and this limits the usefulness of their content for the purpose of environmental monitoring. Monitoring programs require accurate and detailed geo-temporal stamps so that observations can be attributed to, and linked with other environmental data. As summarised in Steiger et al. (2015) progress is being made in identifying geographic attributes associated with tweets, including Density-Based Spatial Clustering. Another avenue would be to identify what the post is talking about to then derive where the post was sent from.

Analysis of the tweeter profiles revealed that the majority of data providers were domestic and international visitors to the region. This is important for two reasons. First, analysis of Twitter data for the purpose of environmental monitoring relies on a critical mass of tweets. If the local population is small the system then relies on a sufficient number of other people visiting the region and sharing their observations. In other words, natural environments in remote locations that are not visited by tourists are less suited for a social media based monitoring system. Second, research has shown that travellers are considerably more likely to share information through their online networks than those staying at home (Travelmail Reporter, 2013). Thus, again, the potential significance of tourists as generators of collective knowledge is implied. The dominant language of tweets in this research was English, but future analysis of other platforms (e.g. Chinese social media such as Weibo) could focus on other languages to increase the data volume and broaden the generality of findings. When using information provided by tourists it is important to understand visitors' different cultural and geographic contexts, and consider previous experiences or comparisons they may draw between the GBR and other marine environments. Thus, information is highly subjective and to be interpreted through the lens of the respective data provider. This in itself introduces a range of factors that make the integration of these soft data with 'hard scientific monitoring' data challenging.

Thus, using social media information to monitor the environment accepts a people-focused approach. Therefore, data only provide insights into what matters to people. In this research it was found that people share special moments and unexpected encounters whilst diving or swimming in the waters of the GBR. If the experience of the Reef was 'normal', it is possible that people chose not to provide this information. This confirms earlier research by Cassa et al. (2013) who found that people engage in social media when the event/situation is outside the norm. Unusual weather events, for example, are likely to feature more often on social media than expected weather (Hyvarinen and Saltikoff, 2010), adding useful information to regular meteorological monitoring for situations that 'count'. This information bias has several implications. First, events discussed on social media are not representative, and second, analysis of 'normal conditions' is likely to be less fruitful. Possibly because of this reason, the monitoring of (slow) environmental change has been less prominent in social media analysis than extreme events. The opportunity, however, then lies in detecting substantial changes or disasters, for example an oil spill.

Whilst the volume of data from 208,525 tweets initially seemed large – and useful for monitoring socio-economic trends or visitor satisfaction – the actual number of tweets discussing the Reef and the surrounding marine environment was much smaller. Geo-tagged tweets posted from the defined GBR spatial polygon could be complemented through API streaming of tweets posted from anywhere in the world and containing words of interest. This approach may capture relevant tweets where the GPS is disabled but which were actually posted from GBR region. However, to correctly identify these tweets it is necessary to implement machine learning methods. The current volume of geo-tagged tweets is comparable with numbers used in related research on environmental monitoring (Daume, 2016). For example, tweets that contained important keywords ranged between 307 for coral and 804 for fish.

The relevant tweets were typically non-observational, which means that their ecological content was unspecific (Daume, 2016). Over time, and as emphasised by Avvanuti et al. (2016) it is critical to improve the filtering process to capture relevant tweets (Shook and Turner, 2016). Machine learning can assist in refining filters over time. In the meantime, relatively small numbers of tweets in this current sample means that it is not robust to disaggregate information in space and time. As a result, the monitoring of people's perception of the Reef environment can be, at best, at an aggregate level. As a socially moderated biophysical data input into the existing biophysical monitoring program of GBRMPA's the Twitter feeds are therefore of limited value. In the future, social media users could be actively encouraged to provide information, for example through a designated hashtag. Such a system is successfully implemented in New South Wales to monitor fire hazards, where citizens supply location-specific information using the hashtag #nswfires (Vivacqua and Borges, 2012).

In addition to the environmental interests, the information could be useful for other purposes, for example the monitoring of visitor experiences and changes in perceived aesthetic value. The sentiment analysis presented here showed that a larger number of tweets are positive than negative. In the absence of a benchmark this insight needs to be treated with caution, but monitoring over time might reveal improvements or declines in sentiment associated with key aspects of the GBR region. Advances in the sentiment algorithm will improve the accuracy of results and possibly identify a larger share of polarised tweets, compared with neutral ones. Detailed content analysis and manual annotation of relevant tweets identified in this sample is an important next step to further assess the usefulness of the information provided (Daume, 2016). Ultimately, the idea is that significant changes in sentiment for particular aspects of the GBR could be systematically related to changes in the environmental quality, for example coral cover, water clarity or species diversity.

Despite some of the shortcomings, the research provides an important basis for further exploration of the use of social media data for environmental monitoring. Next steps involve the integration of social data with biophysical data (e.g. water quality measurements, meteorological data) and the use of imagery. Hyvarinen and Saltikoff (2010), for example, found that analysis of photos uploaded on Flickr was relatively accurate when compared with actual weather data. Analysing imagery presents a wide range of new opportunities, including visual changes to sites, for example as a result of coastal erosion, littering or other environmental impacts. Hybrid approaches, where collective sensing is combined with citizen science and expert monitoring, might present an interesting avenue for the GBRMPA (Connors et al., 2012), with the human sensing data consisting of both text and photo content. Whilst the current volume and specificity of Twitter-based data might be insufficient, improvements in filter, proactive engagement of potential information providers and extension to other platforms may well deliver the data required to build a comprehensive and integrated data management system (Avvanuti et al., 2016).

6. Conclusion

This research sought to capitalise on the fast growing availability of mobile Internet data produced by human sensors. More specifically, information provided through the social media platform, Twitter, was used to assess the potential of mining such data for the purpose of complementing traditional environmental monitoring. The Great Barrier Reef in Australia served as a suitable case study region. It was found that the geographic spread of Twitter posts is, at least in theory, sufficient to gather 'collective knowledge' provided by those enjoying the Reef at its coastal fringes, islands or water activity related locations. The analysis of tweets showed that the data volume is greatly enhanced by tourists visiting the region, highlighting the important of non-resident sensors in developing these novel citizen science-like approaches. The keyword and sentiment analyses highlighted that the actual number of marine environment related tweets are small and would need to be boosted through various mechanisms (e.g. a designated hashtag). Monitoring positive or negative sentiments in relation to key aspects of the marine environment, however, seems a promising avenue to track change, especially when these social perceptions are then integrated with biophysical data to identify patterns of correlation. Whilst unearthing some of the challenges associated with using human sensor data, this research has demonstrated that further exploration of collective sensing for environmental research is worthwhile.

Acknowledgements

This project was funded through a National Environment Science Program (NESP) fund, within the Tropical Water Quality Hub (Project No: 2.3.2). We would also like to thank key stakeholders for the feedback and input at several outreach seminars.

References

Addison, P., Walshe, T., Sweatman, H., Jonker, M., Anthony, K., MacNeil, A., Thompson, A., Logan, M., 2015. Towards an Integrated Monitoring Program: Identifying Indicators and Existing Monitoring Programs to Effectively Evaluate the Long Term Sustainability Plan. Australian Institute of Marine Science, Townsville, Queensland.

Alaei, A.R., Becken, S., Stantic, B., 2017. Sentiment analysis in tourism: beginning of a new research paradigm? Second revision in J. Travel Res. Forthcoming.

Australian Government, Queensland Government, 2015. Reef 2050 long-term sustainability plan. Available (19/12/16). https://www.environment.gov.au/marine/gbr/publications/reef-2050-long-term-sustainability-plan.

Avvanuti, M., Cimino, M.G.C.A., Cresci, S., Marchetti, A., Tesconi, M., 2016. A framework for detecting unfolding emergencies using humans as sensors. SpringerPlus. http://dx.doi.org/10.1186/s40064-016-1674-y.

Bohannon, J., 2016. Twitter can predict hurricane damage as well as emergency agencies. Science. Mar. 11, 2016. Available (15/05/16). http://www.sciencemag.org/news/2016/03/Twitter-can-predict-hurricane-damage-well-emergency-agencies.

Brownstein, J.S., Freifeld, C.C., Reis, B.Y., Mandl, K.D., 2008. Surveillance Sans Frontieres: Internet-based emerging infectious disease intelligence and the HealthMap project. PLoS Med. 5 (7), 1019–1024.

Cassa, C.A., Chunara, R., Mandl, K., Brownstein, J.S., 2013. Twitter as a sentinel in emergency situations: lessons from the Boston marathon explosions. PLOS Curr. Disasters. http://dx.doi.org/10.1371/currents.dis.ad70cd1c8bc585e9470046cde334ee4b. July 2, Edt 1.

Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., Ebert, D.S., 2014. Public behavior response analysis in disaster events utilizing visual analytics of microblog data. Comput. Graph. 38, 51–60.

Chaffey, C., 2016. Global social media research summary 2016. Smart Insights. Available (01/11/16). http://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/.

Cheong, M., Lee, V.C.S., 2011. A microblogging-based approach to terrorism informatics: exploration and chronicling civilian sentiment and response to terrorism events via Twitter. Inf. Syst. Front. 13 (1), 45–59.

Commonwealth of Australia, 2015. Reef 2050 integrated monitoring and reporting program. Available (20/11/16). http://www.gbrmpa.gov.au/managing-the-reef/reef-2050/reef-integrated-monitoring-and-reporting-program.

Connors, J.P., Lei, S., Kelly, M., 2012. Citizen science in the age of neogeography: utilizing volunteered geographic information for environmental monitoring. Ann. Assoc. Am. Geogr. 102 (6), 1267–1289.

Couvet, D., Jiguet, F., Julliard, R., Levrel, H., Teyssedre, A., 2008. Enhancing citizen contributions to biodiversity science and public policy. Interdiscip. Sci. Rev. 33, 95–103.

Daume, S., 2016. Mining Twitter to monitor invasive alien species — an analytical framework and sample information topologies. Ecol. Inf. 31, 70–82.

Deloitte Access Economics, 2013. Economic Contribution ofthe Great Barrier Reef. GBRMPA, Townsville.

GBRMPA, 2017. Significant coral decline and habitat loss on the Great Barrier Reef. Available (09/07/17). http://www.gbrmpa.gov.au/media-room/latest-news/coral-bleaching/2017/significant-coral-decline-and-habitat-loss-on-the-great-barrier-reef.

Great Barrier Reef Water Science Taskforce and Department of Environment & Heritage Protection, 2016. Great Barrier Reef Water Science Taskforce: Clean Water for a Healthy Reef – Final Report. Queensland Government. Available (20/03/17). http://www.ehp.qld.gov.au/assets/documents/reef/gbrwst-finalreport-2016.pdf.

Gharesifard, M., Wehn, U., van der Zaag, P., 2017. Towards benchmarking citizen observatories: features and functioning of online amateur weather networks.

J. Environ. Manag. http://dx.doi.org/10.1016/j.jenvman.2017.02.003.
Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C., 2014. Geo-located Twitter as proxy for global mobility patterns. Cartogr. Geogr. Inf. Sci. 41 (3), 260–271.
Hutto, C., Gilbert, E., 2014. VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media.
Hyvarinen, O., Saltikoff, E., 2010. Social media as a source of meteorological observations. Mon. Weather Rev. 138, 3175–3184.
Keeler, B.L., Wood, S.A., Polasky, S., Kling, C., Filstrup, C.T., Downing, J.A., 2015. Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes. Front. Ecol. Environ. 13 (2), 76–81.
Kirilenko, A.P., Molodtsova, T., Stepchenkova, S.O., 2015. People as sensors: mass media and local temperature influence climate change discussion on Twitter. Glob. Environ. Change 30, 92–100.
Leung, D., Law, R., Hoof, H. van, Buhalis, D., 2013. Social media in tourism and hospitality: a literature review. J. Travel & Tour. Mark. 30 (1–2), 3–22.
Meyer, R., 2015. How the USGS detects earthquakes using twitter. *The Atlantic*. Available (15/05/16). http://www.theatlantic.com/technology/archive/2015/10/how-the-usgs-detects-earthquakes-using-Twitter/409909/.
O'Leary, D., 2011. The use of social media in the supply chain: survey and extensions. Intell. Syst. Account. Finance Manag. 18 (2–3), 121–144.
Paul, K., Quinn, M.S., Huijser, M.P., Graham, J., Broberg, L., 2014. An evaluation of a citizen science data collection program for recording wildlife observations along a highway. J. Environ. Manag. 139 (15), 180–187.
Rao, A., 2014. Social listening: how market sensing trumps market research. PWC Resil. J. Available (18/11/16) http://www.pwc.com/gx/en/services/advisory/consulting/risk/resilience/publications/social-listening-how-market-sensing-trumps-market-research.html.
Resch, B., 2013. People as sensors and collective sensing-contextual observations complementing geo-sensor network measurements. In: Krisp, J.M. (Ed.), Progress in Location-based Services, Lecture Notes in Geoinformation and Cartography. Springer-Verlag, Berlin Heidelberg, pp. 391–406.
Ribeiro, F.N., Araujo, M., Goncalves, P., Goncalves, M.A., Benevenuto, F., 2016. A benchmark comparison of state-of-the-practice sentiment analysis methods. EPJ Data Sci. 5 (23), 1–29.
Riga, M., Karatzas, K., 2014. Investigating the relationship between social media content and real-time observations for urban air quality and public health. In: Paper Presented at the Proceedings of the 4th International Conference on Web Intelligence. Mining and Semantics (WIMS14), Thessaloniki, Greece.
Schnebele, E., Cervone, G., 2013. Improving remote sensing flood assessment using volunteered geographical data. Nat. Hazards Earth Syst. Sci. 13, 669–677.
Shook, E., Turner, K., 2016. The socio-environmental data explorer (SEDE): a social media-enhanced decision support system to explore risk perception to hazard events. Cartogr. Geogr. Inf. Sci. 43 (5), 427–441.
Simon, T., Goldberg, A., Aharonson-Daniel, L., Leykin, D., Adini, B., 2014. Twitter in the cross fire—the use of social media in the Westgate mall terror attack in Kenya. PLoS One 9 (8), e104136.
Steiger, E., de Albuquerque, J.P., Zipf, A., 2015. An advanced systematic literature review on spatiotemporal analyses of twitter data. Trans. GIS 19 (6), 809–834.
The Nature Conservancy, 2017. Mapping Ocean Wealth. Available (12/05/17). http://oceanwealth.org/.
Tourism Research Australia (2015). Research.Available (20/03/16) https://www.tra.gov.au/research
Travelmail Reporter, 2013. Travel is the main theme among Facebook posts - and we can't stay offline while away. Daily Mail, 3 December. Available (29/01/16). http://www.dailymail.co.uk/travel/article-2517369/Travel-dominant-theme-Facebook-postings–stay-offline-holiday.html.
Twitter, 2016. Twitter developer documentation. Available (10/11/16). https://dev.Twitter.com/streaming/public.
Vivacqua, A.S., Borges, M.R.S., 2012. Taking advantage of collective knowledge in emergency response systems. J. Netw. Comput. Appl. 35 (1), 189–198.