Article:

Wilson, T, Lovelace, R orcid.org/0000-0001-5679-6536 and Evans, AJ orcid.org/0000-0002-3524-1571 (2019) A Path Toward the Use of Trail Users' Tweets to Assess Effectiveness of the Environmental Stewardship Scheme: An Exploratory Analysis of the Pennine Way National Trail. Applied Spatial Analysis and Policy, 12 (1). pp. 71-99. ISSN 1874-463X

https://doi.org/10.1007/s12061-016-9201-7

Figure 1 Click here to download colour figure Figure1.tiff

Figure 2 Click here to download colour figure Figure2.tiff

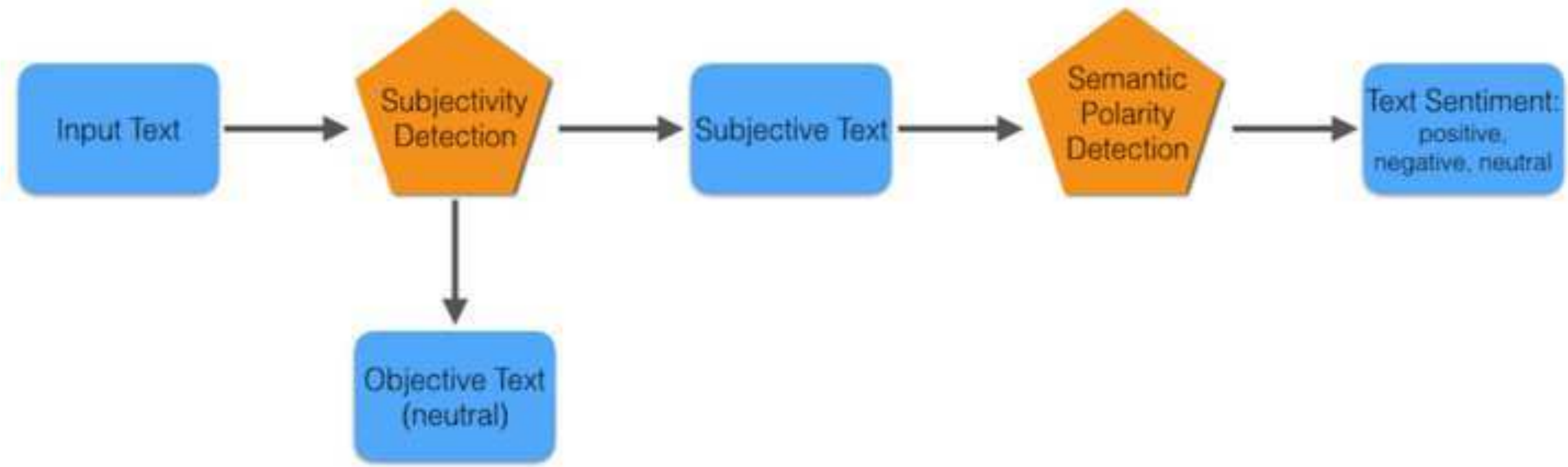


Figure 3

Click here to download colour figure Figure3.tif

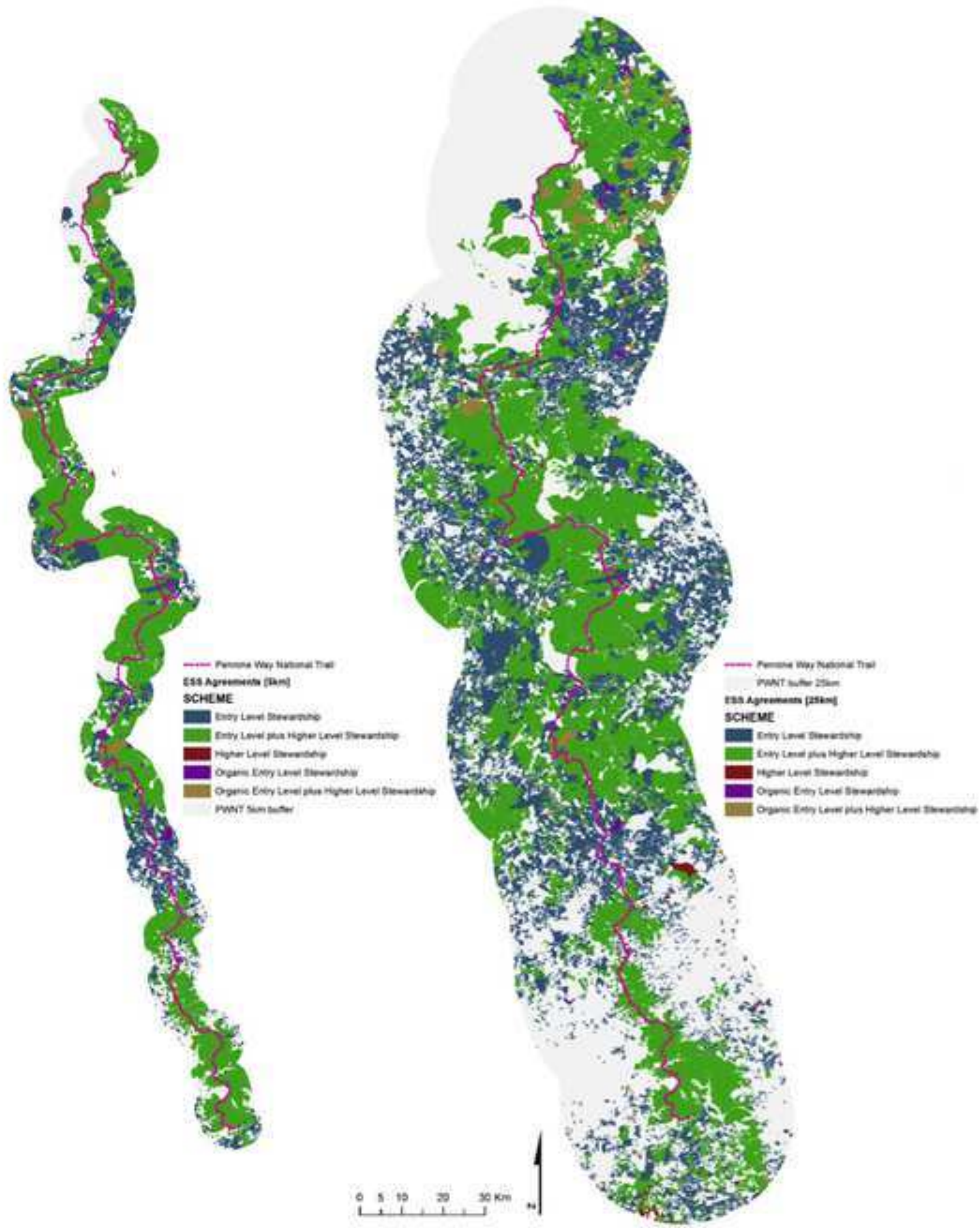


Figure 4

Click here to download colour figure Figure4.tiff

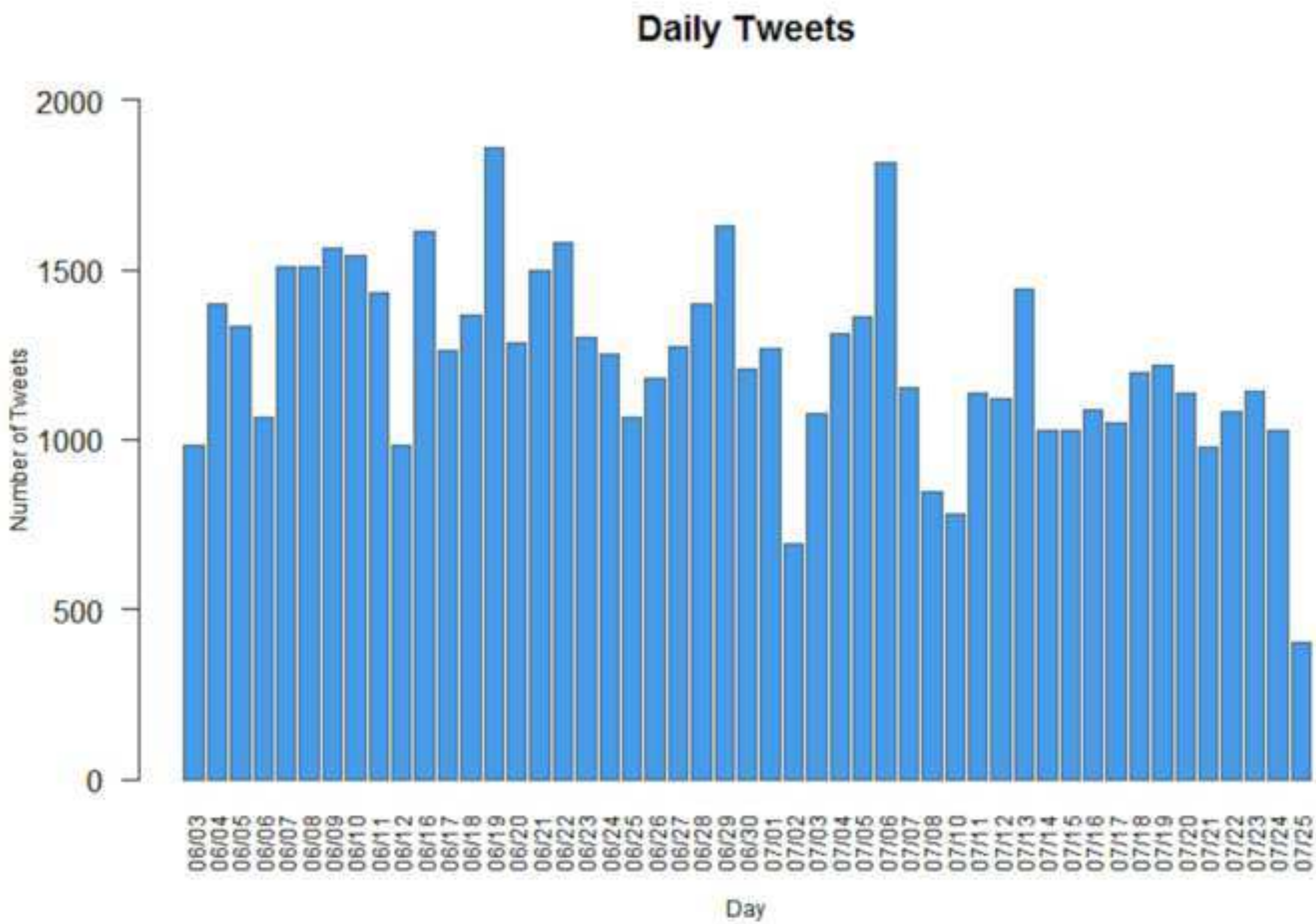


Figure 5

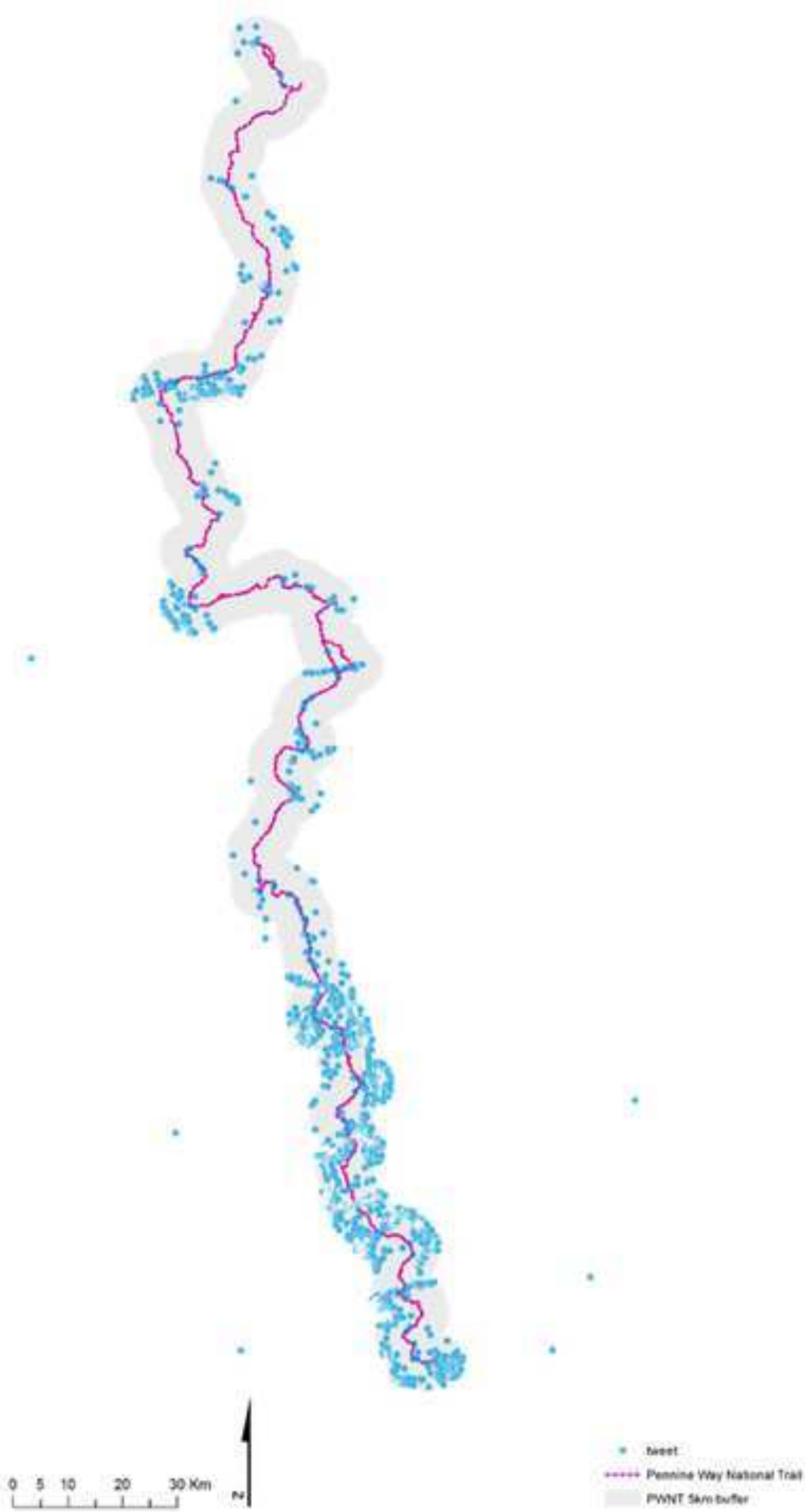

0 5 10 20 30 Km
N
tweet
Pennine Way National Trail
PWNT 5km buffer

Figure 6

Click here to download colour figure Figure6.tif

Figure 7

Click here to download colour figure Figure7.tif

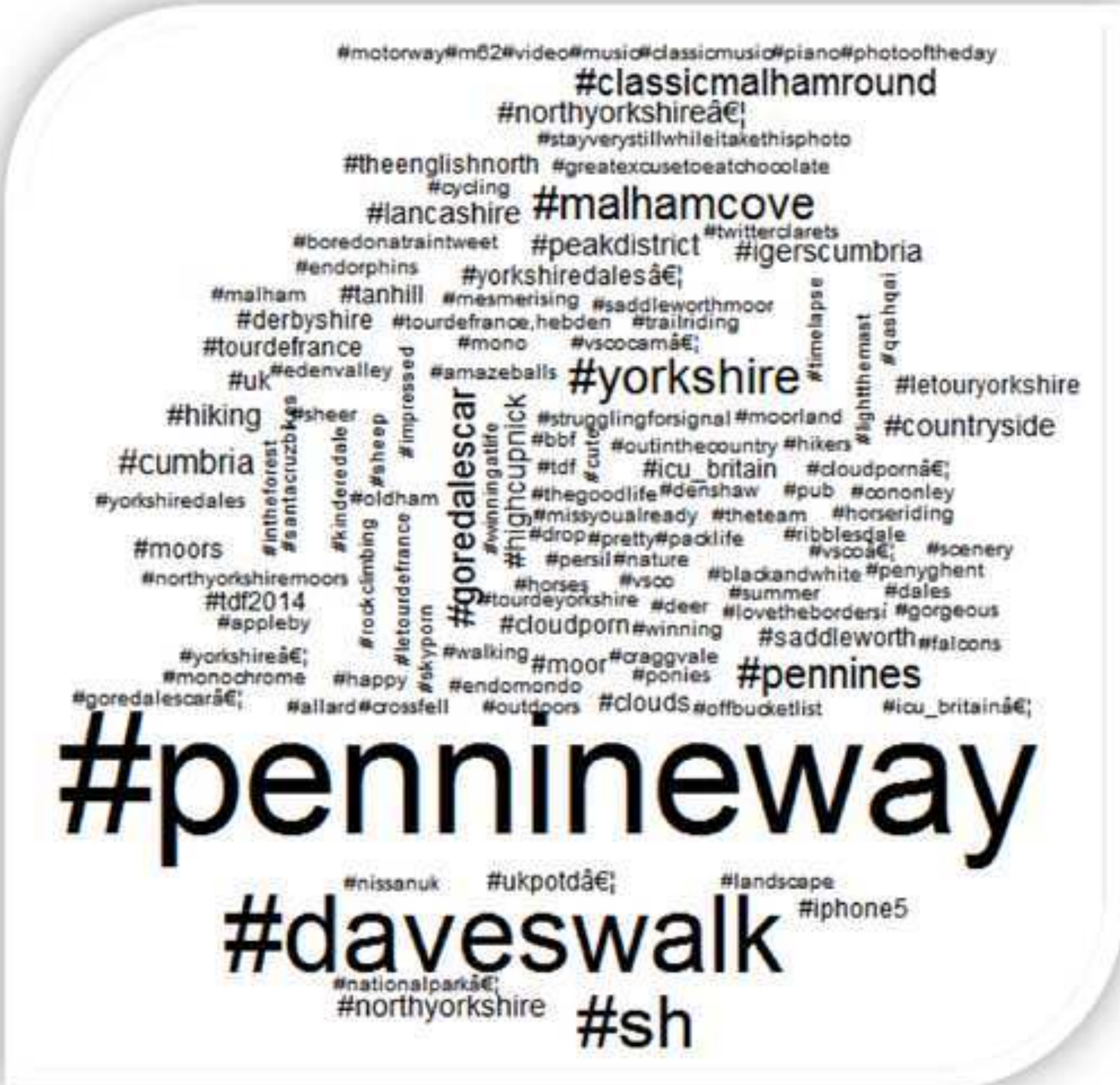


Figure 8

Click here to download colour figure Figure8.tif

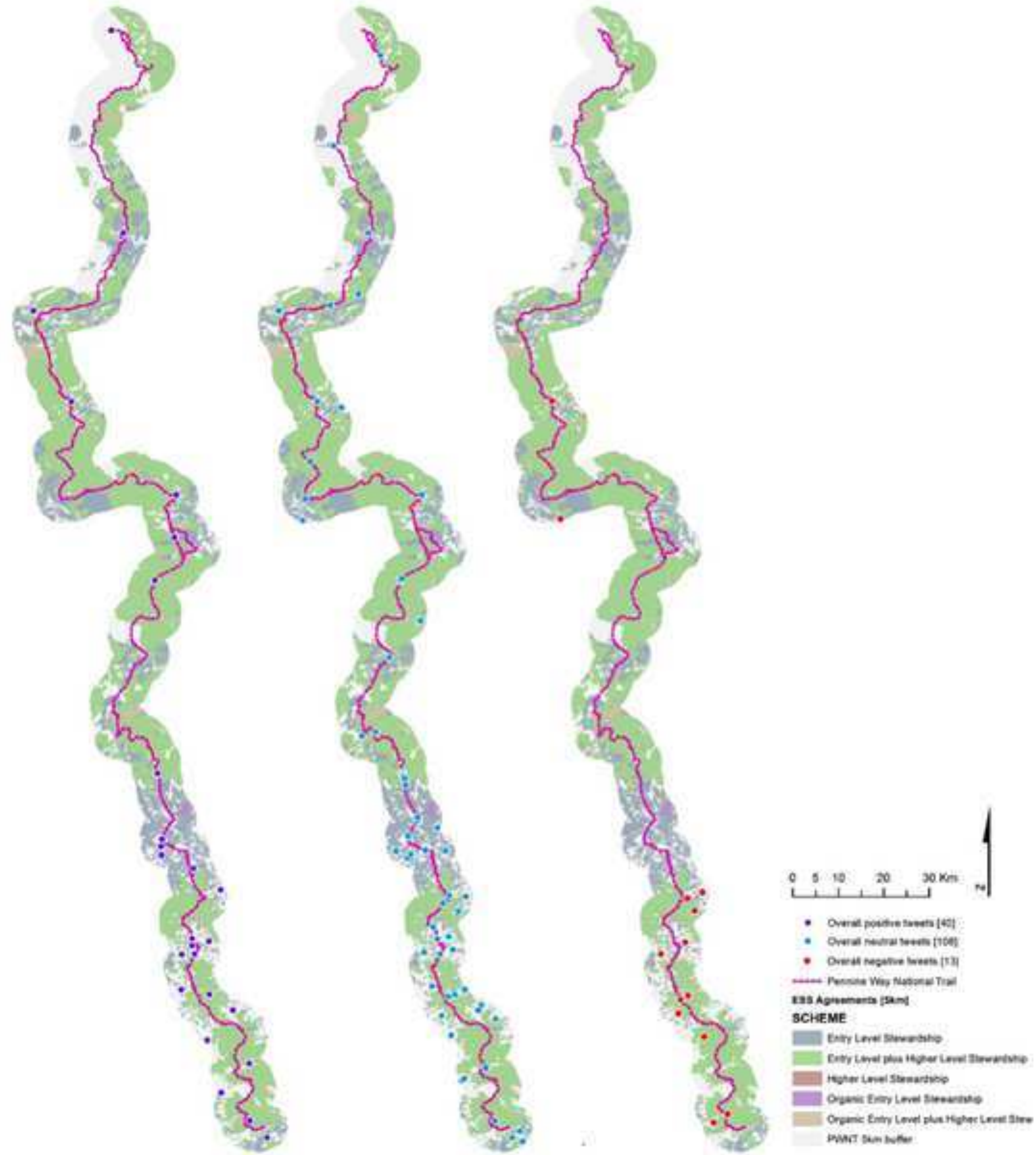
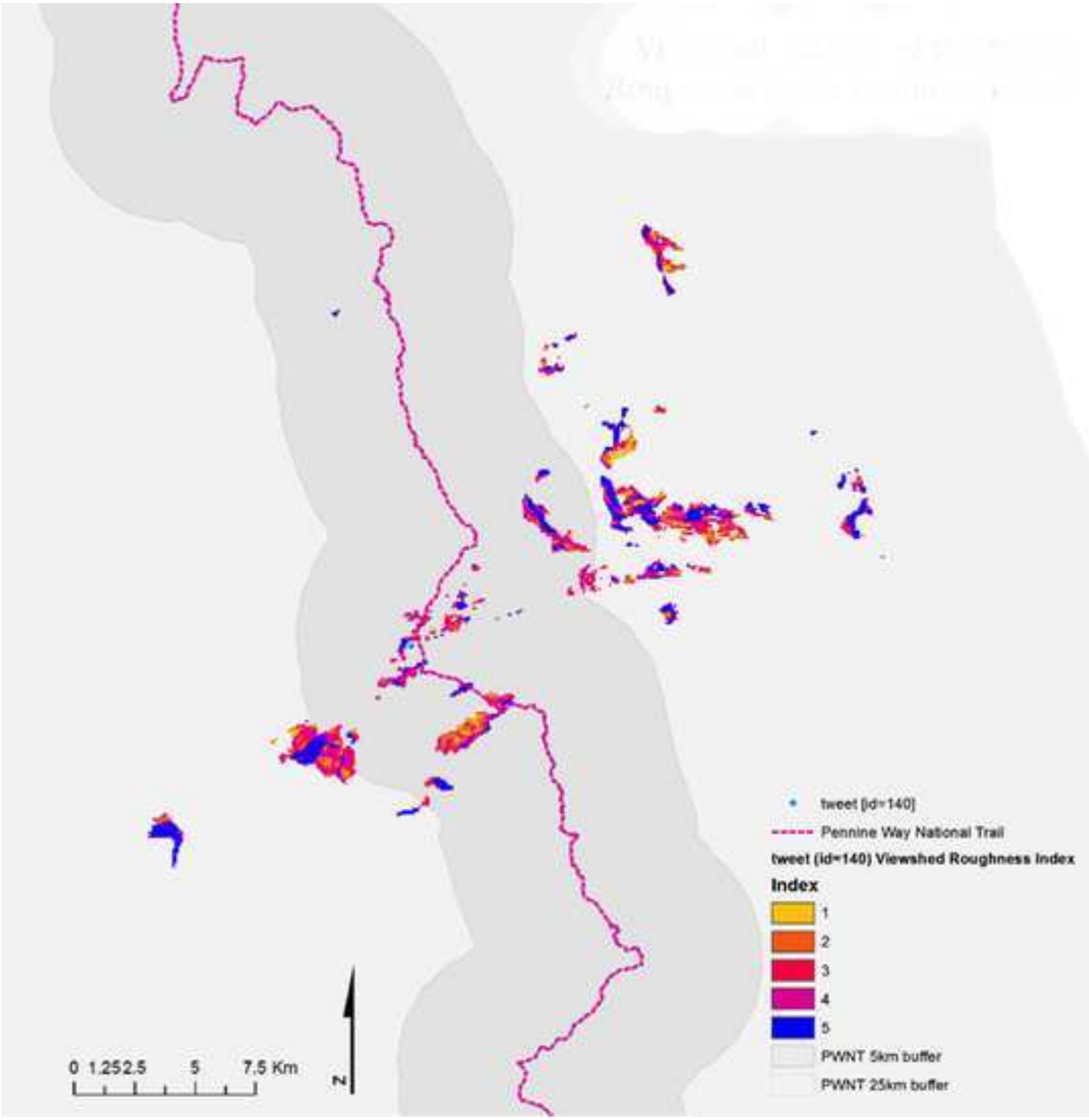


Figure 9

Click here to download colour figure Figure9.tif



A path toward the use of trail users' tweets to assess effectiveness of the Environmental Stewardship Scheme. An exploratory Analysis of the Pennine Way National Trail.

• Author 1[1] • Author 2[1] • Author 3[1]

Abstract

Large and unofficial data sets, for instance those gathered from social media, are increasingly being used in geographical research and explored as decision support tools for policy development. Social media data have the potential to provide new insight into phenomena about which there is little information from conventional sources. Within this context, this paper explores the potential of social media data to evaluate the aesthetic management of landscape. Specifically, this project utilises the perceptions of visitors to the Pennine Way National Trail, which passes through land managed under the Environmental Stewardship Scheme (ESS). The method analyses sentiment in trail users' public Twitter messages (tweets) with the aim of assessing the extent to which the ESS maintains landscape character within the trail corridor. The method demonstrates the importance of filtering social media data to convert it into useful information. After filtering, the results are based on 161 messages directly related to the trail. Although small, this sample illustrates the potential for social media to be used as a cheap and increasingly abundant source of information. We suggest that social media data in this context should be seen as a resource that can complement, rather than replace, conventional data sources such as questionnaires and interviews. Furthermore, we provide guidance on how social media could be effectively used by conservation bodies, such as Natural England, which are charged with the management of areas of environmental value worldwide.

Keywords big data analysis • sentiment analysis • Environmental Stewardship Scheme • Volunteered Geographic Information • National Trails • social media

Introduction

The Environmental Stewardship Scheme (ESS) is an agri-environmental scheme (AES) in England, integrating environmental concerns into the European Commission's Common Agricultural Policy (CAP) (European Commission, 2015). The ESS provides government-financed payments to farmers and land managers in return for a commitment to farming their land with more care for the environment (Smith et al., 2013).

Introduced during 2005 and 2006, a key objective of the ESS is the maintenance and enhancement of landscape quality and character (Natural England, 2011). The ESS was established as a 'broad and shallow' approach to AES in order to extend its reach to high proportions of the countryside (Amy et al., 2013). Therefore the ESS is non-competitive and open to all farmers and land managers whose land is part of the farmed environment and is registered in the Rural Land Register (Natural England, 2013b). With agreements in place on over 70% of agricultural land in England, the ESS represents the country's most widespread approach to environmental management (Defra, 2013).

England's diverse landscapes are connected by a 4000 km network of 15 National Trails which aim to provide greater access to the English countryside (Long Distance Walkers Association, 2014), rewarding natural adventures, and the opportunity for people to be inspired by varied scenery and landscapes (Wood-Gee, 2008). In 2012 approximately 12 million visits were made to England's National Trails (Ramblers, 2012). The maintenance and enhancement of landscape quality and character is one of the primary objectives of the ESS, and evidence suggests that the ESS improves biodiversity (Defra & Natural England, 2008). Based on these observations, one could assume that the ESS can play an important role in providing a positive experience for visitors to National Trails.

Despite the abundant interactions that National Trail users have with England's landscapes and land managed under the ESS, there is currently no method to elicit their opinions regarding the effectiveness of ESS in the maintenance and enhancement of the landscape quality and character. The effectiveness of the ESS tends to be measured in terms of the delivery of environmental benefits and the nationwide penetration of the scheme. The opinions of trail users are generally limited to large-scale surveys, such as the 'Monitor of Engagement with the Natural Environment' (MENE) which broadly examines the adult population's engagement with the natural environment in general (Natural England, 2015a). Previous National Trail User Surveys (The Countryside Agency, 2005; Natural England/Countryside Council for Wales, 2007) have not been conducted since 2007, providing a strong incentive to obtain trail users' opinions using alternative sources.

Author 1: author1@author1.ac.uk

We seek to address this discontinuity in knowledge by exploring the feasibility of using the sentiment conveyed within trail users' public Twitter messages (tweets) to assess the effectiveness of the ESS agreements within the National Trails' corridor. We aim to investigate the feasibility of using trail user sentiment as an assessment of the ESS. As an exploratory analysis, we will not capture the full complexity of landscape quality assessment nor landscape perception. Rather, we will devise a process to extract the sentiment conveyed within trail users' tweets and perform an exploratory analysis that seeks to determine the feasibility of using these perceptions to assess the ESS. The findings of our exploratory analysis will form the basis of recommendations to Natural England regarding the practical feasibility of using social media data as a method of eliciting trail users' opinions of ESS and the National Trail System. Based on this evolving policy and research context the objectives of this paper are to:

- Elucidate a process to select tweets that are relevant to the scope of this study from a larger Twitter dataset;

- Extract and determine the sentiment conveyed within the relevant tweets;

- Conduct exploratory spatial analyses of the geographic origins of the tweets, and their viewsheds;

- Analyse the results of the sentiment and spatial analysis;

- Provide policy recommendations regarding the feasibility of using social media data as a source of trail user feedback.

This study focuses on The Pennine Way National Trail (PWNT). The PWNT travels 431 km (268 miles) along the central upland spine of England connecting the English Midlands to the Scottish Borders (Figure 1) (Walk Unlimited, 2014; Long Distance Walkers Association, 2014). The PWNT was England's first national trail, opening on 24th April 1965 after a 30-year campaign to provide greater access to the English countryside (Long Distance Walkers Association, 2014).74.08% of the land within 5 km of the PWNT is managed under the ESS. Based on trail-counter data collected between 2004 and 2014, an average of approximately 300,000 people visit the PWNT annually (Natural England, 2014b).

Figure 1 Map of England showing the path of the Pennine Way National Trail

Natural England is the executive non-departmental public body of the UK Government that is responsible for management of both the ESS and the National Trail system. Natural England provides oversight and administration of the ESS and interacts with farmers and land managers. To manage the National Trails, Natural England has developed a national framework of guidance, support, and funding, and works with local communities to ensure effective service delivery. New National Trail quality standards have recently been introduced to ensure routes are of the highest standard and connect the finest landscapes. The enhancement of the landscape, natural, and historic features within the trail corridor is a specific quality standard of the National Trail system (Natural England, 2013a). The maintenance and enhancement of landscape quality and character is also one of the key objectives of the ESS.

Background

Environmental Stewardship Scheme

Agri-environmental schemes (AES) integrate environmental concerns into the European Commission's Common Agricultural Policy (European Commission, 2015). AES provide government-funded payments to farmers and land managers in return for a commitment to environmentally sensitive farming practices (Smith et al., 2013). The current working implementation of AES in England is the Environmental Stewardship Scheme (ESS).

AES were first introduced to England in the late 1980's as Environmentally Sensitive Areas, and in the early 1990's as the Countryside Stewardship Scheme (collectively known herein as 'legacy AES'). Legacy AES served as the government's response to increasing levels of agricultural intensification and its negative impacts on wildlife and landscape character (Natural England, 2009a). The aim of legacy AES was to de-incentivise intensive farming (Hodge & Reader, 2010). Although these schemes reduced intensive farming and mitigated the associated negative impacts, they did little to maintain wildlife habitats and landscape features (Natural England, 2009a). As a consequence AES was redeveloped, leading to the introduction of the ESS during 2005 and 2006.

The ESS was established as a multi-objective scheme. The ESS aims to provide the funding and guidance to enable farmers and land managers to fulfill the following five main objectives (Natural England, 2011):

- Conservation of wildlife and their habitats
- Maintenance and enhancement of landscape quality and character
- Protection of the historic environment
- Protection of soils and reduction of water pollution
- Providing opportunities for people to visit and learn about the countryside.

There are two main levels to the ESS: Entry Level Stewardship (ELS) and Higher Level Stewardship (HLS). ELS is the underlying 'broad and shallow' approach (Natural England, 2009a) that is open to all farmers and land managers in England whose land is registered in the Rural Land Register and part of the farmed environment (Natural England, 2013b). HLS is a competitive scheme available to farmers and land managers with land in predetermined areas who demonstrate the ability to provide greater environmental benefits. ELS is a stepping-stone to HLS, it is therefore possible for farms to concurrently have ELS and HLS agreements in place. Organic versions of both ELS and HLS (OELS and OHLS respectively) exist. OELS and OHLS follow the same principles as their respective levels but are available only to organic farms or farms in transition between conventional and organic farming. They also offer a premium payment to reflect the inherent environmental benefits delivered through organic farming (Natural England, 2011).

Entry Level Stewardship

ELS provides a straightforward and flexible approach to environmental management with 65 different management options that allows environmental management to complement farm operations. Each management option has a pre-assigned point value, and ELS is adopted by selecting options that meet an average 30 points per hectare threshold for the whole farm. Adherence to each management option must be sustained for the five-year duration of the agreement to receive the bi-annual payments (Natural England, 2011). ELS management options aim to reduce the intensity of farming to improve the environmental quality of the surrounding area. Some options provide incentives to restore and maintain features that are now redundant in production terms, such as hedges and ditches (Hodge & Reader, 2010), that contribute to the character of the landscape (Natural England, 2014a).

Higher Level Stewardship

Higher Level Stewardship (HLS) is a spatially targeted scheme open to agricultural land in specific areas, pre-designated by Natural England. HLS is a competitive scheme only open to farms that can deliver the greatest level of environmental benefits. HLS agreements are longer term, seeking to deliver significant environmental benefits over 10 years. HLS management options are tailored to the specific features of the farm and the environmental priorities of the surrounding area. HLS can also fund capital work projects on features that contribute to the character of the landscape (Natural England, 2014a), such as hedging, pond creation, or historical building restoration (Natural England, 2011).

It has been argued that the flexibility of the ELS jeopardises the environmental benefits provided, and that the extended reach forgoes spatial targeting (Hodge & Reader, 2010). A prior evaluation of the ESS found that many agreements focused on a limited number of options (Central Science Laboratory, 2007), suggesting that farmers were selecting management options that would involve the least additional work (Defra & Natural England, 2008), or could be achieved at minimal cost (Hodge & Reader, 2010). The existence of 'popular' and 'unpopular' management options leads to gaps in the provision of environmental benefits. Often, management options with high point values prove most popular as they enable farmers to more readily reach their point threshold (Defra & Natural England, 2008). HLS is less affected by these criticisms because entry into HLS is at the discretion of Natural England and dependent on the agricultural land being within a targeted area (Quillerou & Fraser, 2010).

ESS and the National Trail System

England's network of 15 National Trails provides access to the diverse countryside (Natural England, 2013). Given the trails' locations at the heart of the countryside, and the time people may spend on a trail, long distance routes and National Trails provide people with abundant exposure to the countryside and landscapes (Wood-Gee, 2008). The ESS agreements within the corridor of a National Trail therefore have a particularly important role to play in providing positive experiences to trail users. A primary objective of the ESS is the maintenance and enhancement of landscape quality and character. Furthermore, Natural England's quality standards for the National Trails include enhancement of the landscape, natural, and historic features within the trail corridor (Natural England, 2013a). Previous surveys of trail users of England's long distance routes (trails of more than 50 miles in length which includes a majority of the National Trails), found that the primary attraction of National Trails is the quality of the scenery and the landscapes through which the trails pass, and that almost 50% of trail users reported that the landscape was the highlight of their visit (The Countryside Agency, 2005; Wood-Gee, 2008).

Measuring the effectiveness of ESS

Current methods to assess the effectiveness of the ESS generally focus on the environmental benefits provided under the scheme (Franks & Emery, 2013). The report 'Farming and Nature: Agri-environmental schemes in action' (Natural England, 2009b) summarises the key achievements of AES to date which include: halting the deterioration of, and restoring of, priority habitats; increasing populations of scarce farmland birds and bumblebee populations; the maintenance and enhancement of landscape character; the protection of historical features; connecting people to the natural environment; and providing a major contribution to climate change mitigation (Natural England, 2009b).

National Trail User Surveys that were previously used to capture the opinions of trail users (The Countryside Agency, 2005; Natural England/Countryside Council for Wales, 2007) have not been conducted since 2007. The current survey of visitors to the countryside is the 'Monitor of Engagement with the Natural Environment' (MENE) that seeks to examine the adult population's use and enjoyment of, and relationship with, the natural environment (Natural England, 2015a). However, the MENE uses a broad definition of the natural environment, which includes "all green open spaces in and around towns and cities as well as the wider countryside... away from home and private gardens" (Natural England, 2015a p1) and does not focus on specific environmental spaces such as National Trails. The MENE is a quota-sampled, representative sample of the population that has been conducted annually with 46,000-49,000 participants since 2009 (Natural England, 2015a) and costs about £400,000 per year to administer (National Archives, 2014).

Twitter

Twitter is a microblogging site and social media data platform that allows registered users to publically post short messages of up to 140 characters from a computer or mobile device connected to the Internet. These short messages are known as tweets. By default tweets are public, and although it is possible for users to protect their tweets and make them private (boyd et al. 2010), a majority of users do not. Users are also able to select whether their tweets include geographic information that denotes from where the tweet originated geographically. This geographic information is often referred to as a geotag, and is a form of volunteered geographic information (VGI) (Goodchild, 2007). Current estimates suggest that 1-3% of tweets include a geotag (Morstatter et al., 2013; Broniatowski et al., 2013; Hecht & Stephens, 2014).

Twitter is also a type of social network whereby users can follow other users and choose to receive and view their tweets. In contrast with other social media sites, however, this follow relationship need not be reciprocal (boyd et al., 2010). Twitter user profiles are minimal, compared to those of other social networks such as Facebook or LinkedIn, but are public. A previous study of Twitter categorised the main user intentions of using Twitter as daily chatter about everyday life, conversations between users, reporting and disseminating news (Java et al., 2007), and sharing URLs (boyd et al., 2010).

As of September 15th 2015, Twitter is the 9th most visited website in the world, and the 11th most visited website in the United Kingdom (Alexra, 2015). In the fourth quarter of 2014, Twitter had 288 million monthly active users (Statista, 2015). The sheer number of active users and the ability for tweets to be posted from anywhere with access to the Internet means that Twitter generates a constant stream of information.

Given the large volume of information, hashtags have emerged as a method to label and group topics in Twitter. Prefixing a keyword with a '#' symbol generates a hashtag (boyd et al., 2010) which is then included as part of the tweet. The keyword can be anything as chosen by the user, and multiple hashtags can be used. Hashtags are automatically converted to links; clicking hashtags can help others to find tweets of a common theme, or find tweets that are of specific interest to them (Cunha et al., 2011; Kywe et al., 2012).

Twitter provides the functionality to connect via the platform's Application Programming Interface (API) and collect tweets from the Twitter service. Several public APIs exist but it is the Streaming API that allows for the acquisition of tweets in real-time. Although the APIs are subject to constant change, Driscoll and Walker (2014) provide a detailed comparison of the Twitter APIs that were available in 2014.

There is a degree of opacity surrounding the Twitter APIs, and none of the public APIs provide direct, unfettered access to Twitter data. The streaming API, for example, is generally believed to be subject to a 'streaming cap' of about 1% of all tweets at any point in time (Driscoll and Walker, 2014). Furthermore the criteria by which tweets are made accessible to the API is unknown (boyd & Crawford, 2012).

Nevertheless, the accessibility to a constant stream of user-generated social contents has led to a rapid increase in research based upon social media data. As a consequence there has been development of new approaches to the exploration of phenomena (Wilkinson & Thelwall, 2012). To date, Twitter data have been used to study a diverse variety of topics such as real-time event detection during an earthquake (Sakaki et al., 2010), analysis of crisis events such as riots (Proctor, 2013), public sentiment toward a royal birth (Nguyen et al., 2013), and the spread of misinformation and rumour in the wake of a terrorist attack (Starbird et al., 2014).

Sentiment

Sentiment is the view, attitude, or opinion toward an entity such as a situation, event, or item. Knowledge of the sentiment of others is fundamental to the decision-making process, a key influencer of action, and central to human behaviour (Liu, 2012). Sentiment is important to individuals and organisations alike. Organisations require knowledge of how consumers and the general public feel out their products and services (Liu, 2012) in order to remain competitive in the market (Nasukawa & Yi, 2003). Individuals seek the opinions of existing users and consumers before committing to the purchase a good or service (Liu, 2012). In sum there is a human desire to know what other people think (Pang & Lee, 2008).

Prior to the growth of the Internet, businesses traditionally obtained the sentiment held by consumers and the general public through the expensive deployment of customer satisfaction surveys and the convening of user groups (Nasukawa & Yi, 2012; Liu, 2012). At the individual level, people would simply ask friends and colleagues for their recommendations. Both of these still hold true, but the rise of the Internet and mobile computing means that people are able to publically review products and services, through dedicated review sites such as Yelp! and TripAdvisor, or via social media sites such as Twitter and Facebook. As a consequence sentiment has established itself as a public commodity and individuals and businesses alike are able to obtain and act upon the opinions and experiences of those they may not personally know, and who are unlikely professional reviewers (Pang & Lee, 2008).

Sentiment Analysis

Sentiment Analysis (SA) is the research area within the field of Natural Language Processing (NLP) concerned with the study of sentiment, opinions, evaluations, attitudes, and emotions towards a subject (Thelwall et al., 2012, Liu, 2012), often extracted from text. Initial research in SA was driven by commercial interests that sought to understand the role of sentiment in consumers' decision-making. Much in the same way that the internet, mobile computing, and the social media have shaped the way in which sentiment is publically shared, the technologies have also been responsible for renewed interest in SA, in particular SA of social media data, such as that found on Twitter. For example, SA of social media data have been used in the field of social research (e.g. Thelwall et al., 2010; Thelwall et al., 2012; Thelwall et al., 2011; Go et al., 2009), to gain insights into particular events (Thelwall et al., 2011), and to study the affective dimension of the social web (Thelwall et al., 2012).

In its most basic form, SA describes the task of using NLP and text linguistics to automatically identify the semantic polarity of text through identification of the positive and negative opinions that are expressed within that text (Thelwall et al., 2010; Pang & Lee, 2008). SA is computationally challenging because sentiment can be expressed in subtle, nuanced ways, generally in opposing classes (e.g. positive or negative), or with conformity to a numerical scale (e.g. a 1 to 5 star rating) (Pang & Lee, 2008). The process of SA is a multi-stage process of, at minimum, subjectivity detection and semantic polarity determination (Figure 2). Subjectivity detection determines whether a text, sentence, word, or feature is subjective (i.e. contains sentiment or opinion), or objective (i.e. contains factual information). Semantic polarity establishes whether the identified subjective text, sentence, word, or feature is positive, negative, or neutral (i.e. it conveys no sentiment) (Taboada et al., 2011). Recent developments in SA mean this model has been extended to also include sentiment strength detection, which measures the strength of sentiment in a text, and multiple sentiment detection, which aims to detect the range of emotions that may be present in a given text (Thelwall, 2013a). SA of social media data presents additional challenges for sentiment analysis algorithms due to poor language use, deliberate non-standard spellings, abbreviations of words, and the use of emoticons (Thelwall et al., 2010) and emoji -- all inherent characteristics of social media communications. Features such as these may be misinterpreted or missed altogether by sentiment analysis classifiers that have been designed specifically for commercial purposes (Thelwall et al., 2012). To overcome some of these challenges, SentiStrength is a SA tool that has been specifically developed for the SA of short, informal social media text (Thelwall et al., 2011, 2012, 2013a). In this study we chose to utilise the SentiStregnth tool (described in detail below) rather than develop our own NLP algorithm to conduct SA.

Figure 2 A flow diagram to illustrate the sentiment analysis process

SentiStrength

SentiStrength is a SA tool that has been specifically developed for the SA of short, informal social media text (Thelwall et al., 2011, 2012, 2013a). SentiStrength is a computer algorithm that uses lexicons (dictionaries) annotated with the semantic orientation of words, which are referenced to calculate the semantic orientation of a given text (for an example see Turney, 2002) (Taboada et al., 2011, Thelwall et al., 2012). SentiStrength uses widely available lexicons including the Linguistic Inquiry and Word Count program (Pennebaker et al., 2003), the General Inquiry list of sentiment terms (Stone et al., 1966), and annotations that were made during the development of the tool (Thelwall, 2013). In addition to these core lexicons, SentiStrength references additional lexical lists that optimise it for use with social media data. It is also able to detect the strength of sentiment in text because the lexicon contains human-assigned sentiment strength judgments (Thelwall et al., 2012). These include an emoticon list, an idiom list, a booster word list, a repeated punctuation list, and a negating word list. Each entry in these lists also has an associated strength score assigned to it. The booster word list strengthens or weakens the sentiment-bearing words that follow (e.g. very), the negating word list neutralises sentiment words that follow (e.g. not), and the repeated punctuation list boosts the strength of sentiment words with one or more exclamation points (Thelwall, 2013). All lexicons can be modified by the end user of the tool (Thelwall, 2013).

SentiStrength processes text as follows: The algorithm splits text into unigrams (individual words) and punctuation, and these are then queried against the lexicons (Thelwall, 2013). The algorithm identifies the presence of known sentiment-bearing words and predicts the sentiment of the text based upon the frequency of occurrence of the sentiment-bearing words. Since positive and negative sentiments can coexist within a text (Fox, 2008), SentiStrength returns two integers on a continuous scale; the strength of the positive sentiment (+1 to +5) and the strength of the negative sentiment (-1 to -5) conveyed within the input text. The score +1 and -1 denote an absence of positive and negative sentiment respectively. Therefore, if SentiStrength returns +1 and -1 it implies a lack of overall sentiment within the input text, i.e. the input text is neutral or objective. It is possible to calculate the overall polarity of a text through addition of the two integers (Thelwall et al., 2013).

SentiStrength has been tested on diverse social media data sets and applied to studies in various domains such as a time-series analysis of sentiment expressed on Twitter (Thelwall et al., 2011), a determination of emotional diversity in information dissemination on Twitter (Pfitzner et al., 2012), a sentiment analysis of commute-related smartphone applications in California (New Cities Foundation, 2012), an assessment of the sentiment of short informal text written about celebrities in German (Momtazi, 2012), and a large-scale sentiment analysis of Yahoo! Answers (Kucuktunc et al., 2012).

Data

This pilot study utilised a variety of datasets in order to extract and calculate the sentiment of trail users' tweets, and determine whether this could be used to ascertain the effectiveness of the ESS agreements in place.

The Pennine Way National Trail

A GPX file of the route of the Pennine Way National Trail (Walk Unlimited, 2014) was used to generate both 5 km and 25 km buffers around the trail. The 5 km buffer (5 km PWNT corridor) represents the geographical scope of the project i.e. the area within which trail user's tweets would be collected. The 25 km buffer (25 km PWNT corridor) was used in later viewshed analyses and was chosen to ensure that a 20 km viewshed analysis could be conducted for any tweet, even those toward the edge of the 5 km PWNT corridor. The use of a 20 km viewshed is consistent with previous viewshed analyses of zones of visual impact on protected landscapes, including National Trails, which used a 20 km viewshed based upon consultation with Natural England (Natural England, 2013c).

Environmental Stewardship Scheme Agreements

A shapefile of the ESS agreement boundaries in England (Natural England, 2014e) was clipped to the extent of the 25 km and 5 km PWNT corridors. Non-spatial data for each individual agreement included the level of the agreement (e.g. ELS, HLS), and details about the farm, the duration of the agreement, etc. Figure 3 illustrates the spatial distribution of ESS agreements within the PWNT corridors.

Figure 3 The spatial distribution of ESS agreements within the 5km [left] and 25km [right] PWNT corridors. © Natural England copyright. Contains Ordnance Survey data © Crown copyright and database right

Twitter data

The Twitter data for this research was acquired through a Twitter data-harvesting project conducted at the University of Leeds (Lovelace, 2014). The dataset for this paper was provided in the form of a comma separated file with each row of the file representing a single instance of a tweet. Aside from the text of the tweet (TweetText), the dataset also included additional metadata about each tweet:

- a unique id (TweetID)
- date the tweet was created (DateCreated)
- time the tweet was created (TimeCreated)
- the number of followers of the sender (n_followers)
- the number of others the sender follows (n_following)
- the total number of tweets sent by the sender (n_tweets)
- the sender's location (user_location). This refers to the sender's self-disclosed location from their profile, not the user's location at the time of sending the tweet.

- The location from which the tweet originated was provided by the geotag fields in the dataset; longitude and latitude. Every tweet in the dataset included this geocoded information, which represents approximately 1-3% of all tweets (Morstatter *et al.*, 2013; Broniatowski *et al.*, 2013; Hecht & Stephens, 2014).

The Twitter dataset represented 49 days of data collection between 2014-06-03 and 2014-07-25 inclusive, a total of 60,466 geotagged tweets and their associated metadata (no tweets were collected on 2014-06-13 to 2014-06-15 inclusive, or on 2014-07-09). The maximum number of daily tweets sent was 1860 (2014-06-19), the minimum for a full day was 692 (2014-07-02), and the mean was 1,234 (Figure 4). 1297 tweets (2.15%) contained one or more hashtags. In total there were 15,090 hashtags within the tweets, 7,866 of which were unique. 9785 tweets (16.18%) contained one or more URLs to an external source or photo. Figure 5 illustrates the spatial distribution of the tweets in the Twitter dataset prior to further processing.

Figure 4 The daily frequency of tweets sent for the study period, 2014-06-03 – 2014-07-25

Figure 5 The spatial distribution of 60,466 tweets in the Twitter dataset prior to further processing

Digital Elevation Data

We used 90 m Shuttle Radar Thematic Mapper (SRTM) data (Pope, 2009) as the digital elevation model (DEM) for this research. The SRTM data is a raster dataset with a 90m resolution. Whilst this represents a low resolution product compared to other digital elevation datasets which are available, the 90m product is free and readily available for download. Furthermore, since this work represents an exploratory analysis, the 90m resolution data was chosen as it would allow for quicker processing, and could be exchanged for higher resolution data in subsequent analyses. Figure 6 shows the terrain within the PWNT 25 km corridor.

Figure 6A Digital Elevation Model of the 25 km PWNT corridor, with hillshade

Land Cover Data

We used the 2007 Land Cover Map (LCM) (Morton et al., 2011) to identify land cover within the 25 km PWNT corridor. The 2007 LCM provides land cover information for the United Kingdom. The raster version of the LCM has a 25m spatial resolution, with the value of each pixel representing the most likely broad habitat (of 23 classes of broad habitat) to occur within the area.

Methods

A multi stage process was used to extract the sentiment conveyed within trail users' tweets and determine the effectiveness of ESS agreements. An overview of the steps is provided below, with more detailed description following:

- Spatial selection of tweets based on proximity to PWNT
- Lexical selection of tweets using natural language processing
- TweetText processing (removing duplicate tweets, spurious characters)
- Sentiment Analysis of TweetText using SentiStrength
- Viewshed analyses conducted for each overall positive and overall negative tweet. Viewshed analyses included determination of:
  - viewshed
  - majority land cover class within the viewshed
  - ruggedness within the viewshed
  - presence of ESS agreements within the viewshed.

Spatial and Lexical selection of tweets

The Twitter dataset was spatially clipped to the 5 km PWNT corridor so as only to include tweets that originated within 5 km of the PWNT. The identification of tweets relevant to PWNT use was then completed through lexical selection. Lexical selection involved searching the TweetText of each tweet using case-insensitive regular expression terms. An approach of trial and error was used to ascertain the search terms that returned relevant results. Table 1 is a list of the 20 search terms that were used in the final selection.

pennineway	pennine	moors	high force
pennines	hike	kinder scout	cauldron
hiking	path	stoodley pike	high cup nick
rambling	pen y ghent	top withins	cross fell
moor	tan hill	malham cove	cheviot

Table 1 The final search terms used to select tweets relevant to PWNT use

The results for each of the selections were combined into a single dataset. Since some tweets contained multiple search terms the results were searched and purged of duplicate tweets to prevent introduction of bias into the results.

Some tweets originated from within the 5 km PWNT corridor, and contained relevant search terms, but were clearly not relevant to PWNT use. These included traffic reports, broadcasted informational messages, and direct broadcast messages between Twitter users, identifiable as they include '@username'. These tweets were manually removed and not subject to further analysis.

TweetText processing

The tweet selection process resulted in 161 individual tweets, herein referred to as trail users' tweets. The trail users' tweets were those deemed relevant to the project and would be subject to sentiment analysis using SentiStrength. Although we note that it would have been possible to manually analyse the sentiment in the 161 trail users' tweets, we decided to use SentiStrength so that our process could be applied to a larger dataset in the future. Prior to SA some additional processing was required to prepare the data. Although SentiStength includes an emoticon list with emoticon polarities to identify the sentiment of emoticons (Thelwall et al., 2012), it is not able to interpret emoji, which are unicode characters used to convey sentiment (Unicode Inc, 2012). A visual analysis of the trail users' tweets revealed that several tweets contained Unicode character combinations. These Unicode symbol combinations were referenced using online tools (Emojipedia, 2015) to find the meaning and replaced with the text-equivalent sentiment. Finally, spurious characters and excessive whitespace were removed.

Sentiment Analysis

SentiStrength accepts a tab-delimited text file as input. A tab-delimited text file of the TweetID and TweetText of the 161 trail users' tweets were input. Post processing, SentiStrength produced a tab-delimited text file with positive (+1 to +5) and negative (-1 to -5) sentiment scores appended to the end of each tweet. This score is the positive and negative sentiment conveyed within each tweet. A score of +1 or -1 respectively denotes that positive or negative sentiment was not detected. Therefore a tweet with a +1 and -1 score would be treated as neutral (no sentiment conveyed). Although sentiment strength was provided it would not be used in this research. Rather, it was the presence of positive, negative, or neutral sentiment that was of interest. The sum of the positive sentiment and negative sentiment scores was equal to the overall sentiment of the tweet. A positive score denoted positive sentiment and a negative score denoted negative sentiment. A score of 0 denoted no sentiment (neutral)

SentiStrength also provided a duplicate of the TweetText field showing the sentiment score of each individual word. It was noted that two terms, that were specific locations on the PWNT, 'Cross Fell' and 'High Force', produced negative sentiment scores (due to the words 'cross' and 'force' that were negatively scored within the lexicon). This is an example of domain specificity (Thelwall et al., 2012; Thelwall et al., 2011), which describes the requirement to make changes to the lexicon for use in a specific domain. Here, both 'Cross Fell' and 'High force' were added into the idiom list within SentiStrength and given a sentiment score of 0 (no sentiment). After these changes SA was rerun.

The sentiment analysis output was added to ArcMap 10 (ESRI, 2011) and joined by TweetID to the shapefile of tweets for spatial analysis.

Viewshed Analyses

Viewshed analysis describes the computational process of predicting the total area that is visible from a point in space (Kim et al., 2004). Viewshed analysis has a variety of applications including, for example, planning the locations of communication towers (De Floriani et al., 1994) and wind turbines (Kinder et al., 1999), and to identify the impacts of human features on wilderness character (Tricker et al., 2012; Tricker et al., 2013).

In this research we used viewshed analysis to determine the visible areas from each of the overall positive and overall negative tweet locations. Previous research has found that the appeal of National Trails lies within the quality of the scenery and the landscape (Wood-Gee, 2008). The purpose of the viewshed analyses was therefore to establish the experiential qualities of the landscape within the viewshed of the sentiment-bearing tweet locations, and to ascertain whether certain characteristics within the viewshed are consistent with a tweet conveying positive or negative sentiment.

Attributes that characterise the experiential qualities of the landscape have been established in prior work (e.g. Lesslie et al., 1993; Carver, 1996; Carver et al., 2008; 2012). Based upon this research we identified two land attributes that would be used in our viewshed analyses: First, land cover, specifically perceived naturalness; and second, the rugged and challenging nature of the terrain. It has previously been found that striking topographic features and challenging terrain are viewed as qualities of wild land and are therefore associated with providing an inspiring and challenging experience (Carver et al., 2008; 2012). In addition, we added two additional measures; the extent of the viewshed -- the level of visibility from a tweet location, and the extent and type of ESS agreements within the viewshed.

Viewshed analyses were conducted using The Observation Point Tool in ArcMap 10 (ESRI, 2011). The Observation Point Tool allows for the maximum viewshed distance to be set for each point in the dataset. This parameter, RADIUS2, was set to 20 km meaning that the viewshed would only be considered within this distance. This distance was consistent with previous work (Natural England, 2013c), and necessary because defining the maximum distance would benefit the time taken to process the viewshed, which is a computationally intensive process.

Viewshed Calculation

We used ModelBuilder within ArcMap 10 (ESRI, 2011) to calculate the viewshed of each tweet. The model iterated through each of the overall positive and overall negative tweets and calculated the viewshed. We used a binary viewshed model that did not account for distance decay within the viewshed for this exploratory analysis. Therefore, the output for each point was a binary raster dataset that is at the same spatial resolution of the input digital elevation data (90 m). A visible cell was assigned a value 1, and a no-visible cell the value 0. The percentage of visibility was calculated from the output raster based upon the number of raster cells classed as visible and the total number of cells within the >20 km radius of the point. We used a binary viewshed model and did not account for distance decay within the viewshed for this exploratory analysis.

To facilitate further viewshed analyses an input mask was derived from the viewshed of each point. Raster Calculator (ESRI, 2011) was used to convert the non-visible cells from '0' to 'NoData' so that they could be ignored in further calculations. This viewshed input mask would be used to in further analyses to determine the land cover, ruggedness, and the ESS agreements within the viewshed of each tweet.

Land Cover within viewshed

The viewshed input masks and LCM (Morton et al., 2011) were combined using the Raster Calculator (ESRI, 2011) to determine the majority land cover class within each tweet's viewshed. The use of the majority statistic was done with the understanding that it would not capture the full complexity of landscape quality assessment nor landscape perception. However, the majority calculation is computationally inexpensive, and would help us determine the feasibility of using the trail user sentiment to assess the ESS. The land cover within viewshed was calculated by multiplying the viewshed input mask by the LCM (Morton et al., 2011). The result was the categorical land class(es) within the viewshed, and 'NoData' values for areas outside the viewshed. The majority statistic for each output was then calculated. This process was automated through the use of ModelBuilder (ESRI, 2011) to iterate through each of the viewshed input mask datasets and calculate the majority landcover within each.

Ruggedness within viewshed

A Ruggedness Index was derived from the DEM (Pope, 2009) through calculation of the standard deviation of elevation SDelev (Cooley, Unknown; Ascione et al., 2008; Grohmann et al., 2011). SDelev was selected as it can detect regional relief at a variety of scales and is computationally simple to calculate (Grohmann et al., 2011). The calculation of SDelev was done using the ArcMap Focal Statistics tool (ESRI, 2011) with a 3 x 3 cell raster window (equivalent to 270 m on the ground). This moving window passed over the 25 km PWNT DEM.

The resulting output was a raster dataset with float (decimal) values for each cell. This output was reclassified into equally divided quintiles, thus providing relative ruggedness across the 25 km PWNT corridor study area. These values were standardised to a scale of between +1 and +5, with +5 being the most rugged. This provided the Ruggedness Index.

The ruggedness for the viewshed of each overall positive and overall negative tweet was calculated through use of the Raster Calculator (ESRI, 2011). The viewshed input masks and ruggedness index data were used to calculate the ruggedness index within the viewsheds (areas outside the viewshed were assigned 'NoData'). The mean cell roughness of each viewshed was then calculated to provide an average measure of ruggedness within the viewshed.

ESS agreements within viewshed

The 25 km ESS agreement shapefile was converted into a raster dataset with a cell size of 25 m, which was consistent with the LCM (Morton *et al.*, 2011). A smaller cell size was chosen to limit the loss of detail caused when transitioning from a vector to raster format that may affect the boundaries between agreements.

In Raster Calculator, the ESS agreement raster was multiplied by the viewshed input mask. The extent of each ESS agreement type (ELS, HLS etc.) was then calculated as a percentage of the total viewshed for each of the overall positive and overall negative tweets.

Results

Descriptive Statistics

After spatial and lexical filtering the Twitter dataset contained 161 trail user tweets. Using the attributes n_following, n_followers, and n_tweets it determined that the tweets originated from 93 unique users: 87 of the tweets originated from 19 trail users, and the remaining 74 tweets originated from 74 unique trail users. Based upon trail use data for the PWNT it is estimated that there were 63,247 visitors to the PWNT during the 42-day study period (Natural England, 2014b). The trail users' tweets are approximately a 0.25% unrepresentative sample of total trail users for the period.

21 tweets (13.04%) contained one or more hashtags. In total there were 244 hashtags within the trail users' tweets. 132 tweets (81.99%) contained a URL to an external source or photo. The presence of both hashtags and URLs is proportionally greater in the trail users' tweets than in the original Twitter dataset, which was 2.15% and 16.18% respectively. Figure 7 is a word cloud of the hashtags used within trail users' tweets.

29 (18.00%) of the tweets originated from within an ESS agreement: 19 (11.80%) from combination ELS and HLS agreements and 10 (6.20%) from ELS agreements. No tweets originated from land with OELS or OELS and OHLS agreements. The remaining 113 (72%) originated from land not under an ESS agreement.

Figure 7 Word cloud of the hashtags used in trail users' tweets

Trail User Sentiment

SA revealed that 40 tweets were positive overall, 13 negative overall, and 108 contained no sentiment and were classified as neutral. Figure 8 illustrates the spatial distribution of each the overall positive, overall negative, and overall neutral tweets.

Figure 8 The spatial distribution of the overall positive (left), overall neutral (middle), and overall negative (right) tweets. © Natural England copyright. Contains Ordnance Survey data © Crown copyright and database right

Table 2 provides a summary of the sentiment and origin of trail users' tweets, and whether this was from ELS, ELS and HLS, or non-ESS land. As previously mentioned, the majority of tweets originated from land not under an ESS agreement. Tweet origin alone does not offer much insight into the effectiveness of ESS agreements.

Tweet origin (ESS agreement type)	Overall Sentiment (counts and percentage)		
	positive	negative	neutral

ELS	1	2.50%	0	0%	9	8.33%
ELS + HLS	6	15.00%	2	15.38%	16	14.81%
Non-ESS	33	82.5%	11	84.62%	83	76.86%
Total	40	100%	13	100%	108	100%

Table 2 The origin of each of the tweets.

Further investigation of the 105 neutral tweets revealed that 94 (89.52%) contained a URL within the TweetText. This is proportionally higher than in the trail users' tweets before sentiment analysis (81.99%), and the Twitter data before processing (16.18%). Sharing URLs has been identified as a significant aspect of Twitter use (boyd et al., 2010). Aside from pointing to external websites and news sources URLs are also the method by which images are embedded within a tweet. Further assessment of the neutral tweets confirmed this: Of the 94 tweets containing URLs, a majority pointed to images on Twitter (44) and Instagram (28). 10 URLs pointed to other external sites including 'check-ins' at a location (5), an activity tracking application (1), and external websites (4). The remaining 12 links, although broken, pointed to Twitter (3) and Instagram (9).

Viewshed Analyses

53 viewshed analyses were conducted for each of the overall positive and overall negative tweets (positive n=40, negative n=13). Due to the volume of maps not all are included in this report. Figure 9 is an example of the viewshed analysis for the ruggedness index of a specific positive tweet (TweetID=140). Table 3 summarises the viewshed analyses results each of the overall positive and overall negative tweets.

	Viewshed Analyses	Overall Positive Tweets [n=40](mean values)	Overall Negative Tweets [n=13](mean values)
	% Visibility from tweet location	2.19	1.73
Within tweet viewshed	Majority Land Cover Class	4	4
	Total % of Majority Land Cover Class	42.13	36.37
	Ruggedness Index	1.47	1.70
	% ELS	26.61	13.07
	% ELS + HLS	37.31	41.00

	% HLS	0.21	1.33
	% OELS	0.37	0.22
	% OELS + OHLS	0.71	0.04
	% Total ESS	65.22	55.66

Table 3 Summary of results for of the viewshed analyses for each of the overall positive and overall negative tweets

Figure 9 An example of a viewshed analysis showing the ruggedness index within the viewshed of a positive tweet (TweetID=140)

A one-way between subjects ANOVA was conducted to compare the effect of each of the viewshed analyses (percentage visibility within viewshed, percentage of majority landcover within viewshed, ruggedness within viewshed, and percentage of ESS agreements within viewshed) on tweet sentiment. For each of the viewshed attributes the test was run on values above or equal to the mean, and values below mean.

There was not a significant link between percentage visibility and trail user sentiment at the $p<.10$ level for the three conditions [$F(1, 51) = 1.56$, $p = 0.22$].

There was not a significant link between percentage majority landcover and trail user sentiment at the $p<.10$ level for the three conditions [$F(1, 51) = 0.64$, $p = 0.43$].

There was not a significant link between ruggedness and trail user sentiment at the $p<.10$ level for the three conditions [$F(1, 51) = 2.34$, $p = 0.13$]. However, this result does imply that ruggedness within the viewshed might be shown to have a potential impact on trail user sentiment if more data were collected and the test strengthened.

There was a significant relationship between percentage of ESS agreements within the viewshed and trail user sentiment at the $p<.10$ level for the three conditions [$F(1, 51) = 4.15$, $p = 0.05$]. This result suggests that the proportion of ESS agreements in place within the viewshed is potentially related to trail user sentiment but more data are required to .

Discussion

As an exploratory analysis this paper has demonstrated a process to select tweets that are relevant to the scope of the study from a larger Twitter dataset, and extracted the sentiment conveyed by trail users.

After the tweet selection process only 161 tweets that could be subject to spatial and viewshed analyses remained. This reflects the short collection period of 42 days. Nevertheless, even this dataset has provided practical results, and gives confidence that a longer study in conjunction with the organisations involved would yield further insights. The origin of a majority of these tweets was from land not managed under ESS, yet use of the viewshed as the unit of analysis allows reflection on the ESS management process. The viewshed analyses do imply possible impact on trail user sentiment, and therefore warrant further study. The trend between ruggedness and sentiment, while not significant demands additional investigation, not least because, if real, does appear to show that sentiment may be positive when the land is less rugged. This would be contrary to previous studies that found ruggedness of the landscape signifies the wild character and challenging nature of the terrain, which is generally valued (Carver et al., 2008). The weakness of the association (and the potentially subsequent negative relationship) may be because of the ruggedness metric used, which takes the mean value across the viewshed, and is thus unlikely to capture the full complexity of landscape quality assessment. Nevertheless, the findings suggest that landscape quality assessment and landscape perception warrant additional investigation with a suite of alternative metrics that seek to capture the complex and nuanced properties.

More significantly for this study, the percentage of ESS agreements within the viewshed was shown to have a relationship with trail user sentiment. Relatively high proportions of ESS agreements within the viewshed (greater than the mean value) correlate to more positive tweets. This may be because the management plans positively affect land preservation, or it may be because the allocation of ESS agreements has been passively biased to areas that demanded the preservation of well-loved aesthetic qualities. However, since ELS is not spatially targeted agreements are prevalent across the country, the relationship between trail user sentiment and proportion of ESS agreements within the viewshed does warrant further investigation. What is certain is that is viewsheds containing more areas that are not under the ESS management have lower satisfaction, and this should give the organisations involved some indication that the schemes are either actively working, or they have passively covered the right landscapes. As new agreements come into being, this methodology gives the opportunity for assessing the worth of plans from a combination of these two perspectives: an increase in satisfaction would indicate an active contribution by the scheme to landscape satisfaction, while a decrease would indicate that passive (or active) protection has not worked.

Assessment of the neutral tweets revealed that a very high proportion, close to 90%, contain a URL, which could be a link to an image. Work by Borth et al. (2013) found that tweets conveying the sentiment visually (through images) could be characterised by the short length of tweet text. Although the Borth et al. (2013) study does not provide details of the sentiment of the short text, it does present an interesting avenue for further research of the images attached to trail users' tweets, and also images posted to other photo-sharing websites such as Flickr (Yahoo, 2016) and Geograph.org (Geograph Project Limited, 2016).

Another factor that should be acknowledged is the role of the weather on trail user perceptions. The UK's weather is renowned for its unpredictability, and the Pennine Way is often remote and exposed. The weather will determine the visibility, but also perhaps affect trail user perceptions of the visible areas.

Issues with social media data

There are several issues with the use of social media data that deserve attention. First is accessibility to the data. As previously identified, none of Twitter's public APIs provide direct, unfettered access to Twitter data, rather the Twitter API is believed to be subject to a 'streaming cap' of about 1% of all tweets at any point in time (Driscoll & Walker, 2014). In reality, it is only the social media companies that have full access to the data (Manovich, 2011), and full control as to who can access the data (boyd & Carwford, 2012). It is therefore important to point out that of any data collected using the public Twitter API there also exists data that is not accessible, and this could be ~99%. Unfortunately, researchers cannot account for this data due to a lack of transparency regarding the exact streaming cap, and the process of selecting which tweets are made available via the API (boyd & Crawford, 2012). Nevertheless, it is important that this is recognised within research.

The second is the representation of the data. Twitter usage requires access to the Internet via a desktop computer or a mobile device. As such, Twitter usage is limited to Internet users. Moreover Twitter usage is not evenly distributed among Internet users (Driscoll and Walker, 2014). Specific to this research, not all trail users are necessarily (mobile) internet users, and even those that are (mobile) internet users may not wish to access social media sites while out hiking. Some trail user's may desire to send tweets while out hiking but are unable to do so due to limitations in the mobile phone network. Furthermore, a person needs both a smartphone and a data plan in order to send a tweet whilst out hiking on the PWNT. Consideration of this is needed to avoid the creation of a 'digital divide' whereby only the opinions of trail users with smart phones and data plans are heard.

Third is the question of ethics. As boyd and Marwick (2011) succinctly put it; *"there is a considerable difference between being in public and being public"* (boyd & Crawford, 2012 p.673). Although Twitter data may be classed as public data, consideration should certainly be given to the subjects of the study. Twitter users should understand that their data is public, unless they specify otherwise in their preferences. Even then, there is a chance that users do not realise the potential use of their social media interactions. In the event that users know their data is public, it still can be the case that they do not intend for their data or tweet to become public (Eckert *et al.*, 2013). In consideration of this, the Twitter dataset in this study did not contain Twitter usernames or personal information, and the contents of the tweets are not published in this paper. If a process as described in this paper is adopted as a method of obtaining the sentiment of trail users' we believe it would be necessary for the agencies involved to disclose the purpose of data collection.

Conclusions and Recommendations

A process to select trail users' tweets from a larger dataset of Twitter and extract the sentiment conveyed about geographical spaces has been developed. The exploratory analysis presented has gone some way to providing an indication of the effectiveness of the ESS with regards landscape satisfaction, but the situation deserves further attention with an enhanced dataset.

As previously mentioned, an interesting avenue of future research is to determine the extent of image sharing in trail users' tweets. Furthermore, are tweeting trail users attempting to convey sentiment through these images? Borth et al. (2013) present research on visual sentiment ontology that could provide a foundation of future research in this area.

Based upon the findings of this research it is recommended that Natural England proactively initiate a social media strategy as a method of eliciting the sentiment of its trail users from their social media data. Provided the campaign does not focus on ESS agreements, the data elicited should be unbiased with regards the nature of the land protection. This research has uncovered that trail users already utilise hashtags within their tweets, and Natural England should select a hashtag with which it would like users to tag their tweets. Assigning a hashtag specific to this campaign would facilitate the grouping and selection of tweets during data analysis. Furthermore, the hashtag can form the basis of a promotional campaign and can be used to encourage its own use.

A social media campaign would provide the opportunity to increase awareness of the use of social media data for representing views to the organisation. This is important from both an ethical and representative perspective: Trail users should be alerted to the fact that the sentiment they convey and comments they make are public. In terms of representation, a greater number of people need to be encouraged to participate in this scheme for it to be made more representative. However, alternative mechanisms will be additionally needed to ensure that trail user opinions are not subject to a digital divide whereby only those with a smart phone and a data plan are able to offer their opinion.

Initiation of a social media campaign is also likely to increase the amount of data available for analyses such as those presented in this report, and allow for the process to be refined further.

Finally, this paper presents a methodology that would be suitable for the analysis of any user sentiment where the management or planning of land use within viewsheds is a key determinant. As such, it is hoped that it will be of use to communities as diverse as civic architects, policing organisations, transport planners, and national park managers.

References

Alexa. (2014). How popular is twitter.com? Available at: http://www.alexa.com/siteinfo/twitter.com. Accessed December 22 2014.

Amy, S.R., Baker, A., Carey, P.D., Dean, H.J., Kirby, V.G., Nisbet, A., Peyton, J.M., Pywell, R.F., Redhead, J.W. & Smart, S.M in Mountford, J.O. & Cooke, A.I. (editors). (2013). Monitoring the outcomes of Higher Level Stewardship: Results of a 3-year agreement monitoring programme. Natural England Commissioned Reports, Number 114.

Ascione, A., Cinque, A., Miccadei, E., Villani, F., & Berti, C. (2008). The Plio-Quaternary uplift of the Apennine chain: new data from the analysis of topography and river valleys in Central Italy. Geomorphology, 102(1), pp.105-118.

Borth, D., Ji, R., Chen, T., Breuel, T., & Chang, S. F. (2013). Large-scale visual sentiment ontology and detectors using adjective noun pairs. Proceedings of the 21st ACM international conference on Multimedia, pp. 223-232.

boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication & Society, 15(5), pp.662-679.

boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In 43rd Hawaii International Conference on System Sciences (HICSS), pp. 1-10.

boyd, d., & Marwick, A. (2011). Social privacy in networked publics: teens' attitudes, practices, and strategies. Paper given at the Oxford Internet Institute. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1925128. Accessed February 13 2015].

Broniatowski, D. A., Paul, M. J., & Dredze, M. (2013). National and local influenza surveillance through twitter: An analysis of the 2012-2013 influenza epidemic. PloS one, 8(12).

Carver, S. (1996). Mapping the wilderness continuum using raster GIS. Raster imagery in Geographic Information Systems. OnWord Press, New Mexico, pp. 283-288.

Carver, S., Comber, L., Fritz, S., McMorran, R., Taylor, S., & Washtell, J. (2008). Wildness Study in the Cairngorms National Park. University of Leeds.

Carver, S., Comber, A., McMorran, R., Nutter, S. (2012). A GIS model for mapping spatial patterns and distribution of wild land in Scotland. Landscape and Urban Planning, 104 (3–4), pp. 395-409.

Central Science Laboratory. (2007). Evaluation of the operation of Environmental Stewardship. Report to Defra MA01028.

Cooley, S. (Unknown). Terrain Roughness, 12 ways. GIS 4 Geomorphology. http://gis4geomorphology.com/roughness-topographic-position/. Accessed February 15 2015.

Cunha, E., Magno, G., Comarela, G., Almeida, V., Gonçalves, M. A., & Benevenuto, F. (2011). Analyzing the dynamic evolution of hashtags on twitter: a language-based approach. In Proceedings of the Workshop on Languages in Social Media, pp. 58-65).

De Floriani, L., Marzano, P., & Puppo, E. (1994). Line-of-sight communication on terrain models. International Journal of Geographical Information Systems, 8(4), pp. 329-342.

Defra. (2013). Developing the potential for Payments for Ecosystem Services: an Action Plan.

Defra & Natural England. (2008) Environmental Stewardship – Review of Progress 2008.

Driscoll, K., & Walker, S. (2014). Big Data, Big Questions| Working Within a Black Box: Transparency in the Collection and Production of Big Twitter Data. International Journal of Communication, 8(20).

Eckert, J., Walker, S., Hemsley, J., Nahon, K., & Mason, B. (2013). Opening the Black Box of Social Media Research Methods: SoMe Ways Forward, Work in progress by the SoMe Lab @ UW. http://somelab.net/wp-content/uploads/2013/01/HICSS_workshop_final29Dec.pdf. Accessed February 13 2015.

Emojipedia. (2015). Emojipedia. http://emojipedia.org/ Accessed February 15 2015.

ESRI. (2011). ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

European Commission. (2015). Agriculture and Rural Development. http://ec.europa.eu/agriculture/envir/measures/index_en.htm. Accessed September 15 2015.

Fox, E. (2008). Emotion science. Basingstoke: Palgrave Macmillan.

Franks, J. R., & Emery, S. B. (2013). Incentivising collaborative conservation: Lessons from existing Environmental Stewardship Scheme options. Land Use Policy, 30(1), pp. 847-862.

Geograph Project Limited. (2016) Geograph.org. Accessed January 15 2016

Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, pp. 1-12.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. GeoJournal, 69(4), pp. 211-221.

Grohmann, C. H., Smith, M. J., & Riccomini, C. (2011). Multiscale analysis of topographic surface roughness in the Midland Valley, Scotland. IEEE Transactions on Geoscience and Remote Sensing, 49(4), 1200-1213.

Hecht, B., & Stephens, M. (2014). A tale of cities: Urban biases in volunteered geographic information. In Proceedings of the International Workshop on Web and Social Media (ICWSM).

Hodge, I., & Reader, M. (2010). The introduction of Entry Level Stewardship in England: extension or dilution in agri-environment policy? Land Use Policy, 27(2), pp. 270-282.

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pp. 56-65.

Kidner, D., Sparkes, A., & Dorey, M. (1999). GIS and wind farm planning. Geographical Information and Planning, pp. 203-223. Springer Berlin Heidelberg.

Kim, Y. H., Rana, S., & Wise, S. (2004). Exploring multiple viewshed analysis using terrain features and optimisation techniques. Computers & Geosciences, 30(9), pp. 1019-1032.

Kucuktunc, O., Cambazoglu, B.B., Weber, I., & Ferhatosmanoglu, H. (2012). A large-scale sentiment analysis for Yahoo! Answers. Proceedings of the 5th ACM International Conference on Web Search and Data Mining.

Kywe, S. M., Hoang, T. A., Lim, E. P., & Zhu, F. (2012). On recommending hashtags in twitter networks. In Social Informatics pp. 337-350. Springer Berlin Heidelberg.

Lesslie, R. (1993). The National Wilderness Inventory: wilderness identification, assessment and monitoring in Australia. In International wilderness allocation, management and research. Proceedings of the 5th World Wilderness Congress, pp. 31-36.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan and Claypool.

Long Distance Walkers Association. (2014). Pennine Way National Trail. http://www.ldwa.org.uk/ldp/members/show_path.php?menu_type=S&path_name=Pennine+Way+National+Trail. Accessed February 25 2015.

Lovelace, R. (2014). "GeoT.csv". Twitter Dataset. [dataset]. University of Leeds. 2014.

Manovich, L. (2011). Trending: the promises and the challenges of big social data.

Momtazi, S. (2012). Fine-grained German Sentiment Analysis on Social Media. In LREC, pp. 1215-1220.

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the sample good enough? Comparing data from twitter's streaming API with Twitter's firehose. arXiv preprint.

Morton, R.D., Rowland, C., Wood, C. Meek, L., Marston, C., Smith, G., Wadsworth, R., Simpson, I.C. (2011). Final report for the LCM2007 - the new UK Land Cover Map. Countryside Survey Technical Report No 11/07 NERC/Centre for Ecology & Hydrology 112pp.(CEH Project Number: C03259).

Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of the 2nd International Conference on Knowledge Capture. pp. 70-77.

National Archives. (2014). Q&A: Monitor of Engagement with the Natural Environment (MENE): The national survey on people and the natural Environment. http://webarchive.nationalarchives.gov.uk/20140711133551/http://www.naturalengland.org.uk/Images/mene-stats-faq_tcm6-37788.pdf. Accessed February 8 2015.

Natural England. (2009a). Agri-environment schemes in England 2009. Report NE194.

Natural England. (2009b). Farming with nature Agri-environment schemes in action. Report NE223.

Natural England. (2011). Looking after your land with Environmental Stewardship. Report NE290.

Natural England. (2013). The best trails in England and Wales. http://www.nationaltrail.co.uk/sites/default/files/best_trails.pdf. Accessed February 25 2015.

Natural England. (2013a). The New Deal: Management of National Trails in England from April. Report NE426.

Natural England. (2013b). Entry Level Stewardship. Report NE349.

Natural England. (2013c). GIS Viewshed Analysis to Identify Zones of Potential Visual Impact on Protected Landscapes. Methodology Report Version 1. http://www.geodata.soton.ac.uk/geodata/viewshed/Viewshed_Analysis_Methodology.pdf. Accessed February 15 2015.

Natural England. (2014a). Assessment of the effects of Environmental Stewardship on landscape character. Natural England Commissioned Report 158.

Natural England. (2014e). Environmental Stewardship Scheme Agreement (boundaries). [shapefile]. http://www.geostore.com/environment-agency/WebStore?xml=environment-agency/xml/ogcDataDownload.xml. Natural England. 2014.

Natural England. (2014b). "Pennine Way Counter Data.csv". Pennine Way Counter Data. [datafile]. Natural England. 2014.

Natural England. (2015a). Monitor of Engagement with the Natural Environment: The national survey on people and the natural environment. Technical Report from the 2013-14 survey. Report JP010.

Natural England/Countryside Council for Wales. (2007). Results of the national trail user survey 2007.

New Cities Foundation (2012), Connected Commuting: Research and Analysis on the New Cities Foundation Task Force in San Jose.

Nguyen, V. D., Varghese, B., & Barker, A. (2013). The royal birth of 2013: Analysing and visualising public sentiment in the uk using twitter. In Big Data, 2013 IEEE International Conference on (pp. 46-54). IEEE.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 1(1-2), pp. 1-135.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. Annual Review of Psychology, 54(1), pp. 547-577.

Pfitzner, R., Garas, A., & Schweitzer, F. (2012). Emotional divergence influences information spreading in Twitter, ICWSM-12.

Pope, A. (2009). UK SRTM DEM. [dataset]. Version 1. EDINA, Edinburgh.

Proctor, R. (2013). How 2.6 m tweets were analysed to understand reaction to the riots. The Guardian. http://www.theguardian.com/uk/2011/dec/07/how-tweets-analysed-understand-riots. Accessed February 15 2015.

Quillerou, E & Fraser, R. (2010). Adverse Selection in the Environmental Stewardship Scheme: Does the Higher Level Stewardship Scheme Reduce Adverse Selection? Journal of Agricultural Economics, 61(2), pp. 369–380.

Ramblers. (2012). National Trails: A fantastic future. http://www.ramblers.org.uk/what-we-do/news/2012/september/securing-a-fantastic-future-for-national-trails.aspx Accessed 7 January 2015.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. In Proceedings of the 19th International Conference on World Wide Web, pp. 851-860.

Smith, S., Rowcroft, P., Everard, M., Couldrick, L., Reed, M., Rogers, H., Quick, T., Eves, C. & White, C. (2013). Payments for Ecosystem Services: A Best Practice Guide. Defra, London.

Starbird, K., Maddock, J., Orand, M., Achterman, P., & Mason, R. M. (2014). Rumors, false flags, and digital vigilantes: Misinformation on Twitter after the 2013 Boston Marathon bombing.

Statista. (2015). Number of monthly active Twitter users worldwide from 1st quarter 2010 to 4th quarter 2014 (in millions). http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/. Accessed February 12 2015.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). The general inquirer: A computer approach to content analysis. Cambridge, MA: The MIT Press.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational Linguistics, 37(2), pp. 267-307.

The Countryside Agency. (2005). Results of the National Trail User Survey 2005. http://www.iprow.co.uk/docs/uploads/UserSurveyReport2005.pdf. Accessed February 23 2015.

Thelwall, M. (2013). Heart and soul: Sentiment strength detection in the social web with sentistrength. Cyberemotions, 1-14.

Thelwall, M. (2013a). Sentiment strength detection for the Social Web. Lecture series at Universitat Rovira i Virgili. http://sentistrength.wlv.ac.uk/documentation/course.html Accessed December 26 2015.

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in twitter events. Journal of the American Society for Information Science and Technology, 62(2), pp. 406-418.

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology, 63(1), pp. 163-173.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 61(12), pp. 2544-2558.

Tricker, J., Landres, P., Chenoweth, J., Hoffman, R., & Scott, R. (2013). Mapping Wilderness Character in Olympic National Park.

Tricker, J., Landres, P., Dingman, S., Callagan, C., Stark, J., Bonstead, L., & Carver, S. (2012). Mapping Wilderness Character in Death Valley National Park. Natural Resource Report NPS/DEVA/NRR-2012/503. National Park Service, Fort Collins, Colorado, 82.

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (ACL), July 6-12, 2002, Philadelphia, PA, 417-424.

Unicode Inc. (2012). The Unicode Consortium. http://www.unicode.org/. Accessed February 20 2015.

Walk Unlimited (2014). The Pennine Way. http://nationaltrail.co.uk/pennine-way Accessed 7 January 2015.

Wilkinson, D., & Thelwall, M. (2012). Trending Twitter topics in English: An international comparison. Journal of the American Society for Information Science and Technology, 63(8), pp. 1631-1646.

Wood-Gee, V. (2008). Long distance recreational routes. Scottish Natural Heritage Commissioned Report No.274 (ROAME No.RO6AA608).

Yahoo. (2016). Flickr.com. Accessed January 12 2016.

A path toward the use of trail users' tweets to assess effectiveness of the Environmental Stewardship Scheme. An exploratory Analysis of the Pennine Way National Trail.

• Tom Wilson[1] • Robin Lovelace[1] • Andrew J. Evans [1]

Abstract

Large and unofficial data sets, for instance those gathered from social media, are increasingly being used in geographical research and explored as decision support tools for policy development. Social media data have the potential to provide new insight into phenomena about which there is little information from conventional sources. Within this context, this paper explores the potential of social media data to evaluate the aesthetic management of landscape. Specifically, this project utilises the perceptions of visitors to the Pennine Way National Trail, which passes through land managed under the Environmental Stewardship Scheme (ESS). The method analyses sentiment in trail users' public Twitter messages (tweets) with the aim of assessing the extent to which the ESS maintains landscape character within the trail corridor. The method demonstrates the importance of filtering social media data to convert it into useful information. After filtering, the results are based on 161 messages directly related to the trail. Although small, this sample illustrates the potential for social media to be used as a cheap and increasingly abundant source of information. We suggest that social media data in this context should be seen as a resource that can complement, rather than replace, conventional data sources such as questionnaires and interviews. Furthermore, we provide guidance on how social media could be effectively used by conservation bodies, such as Natural England, which are charged with the management of areas of environmental value worldwide.

Keywords big data analysis • sentiment analysis • Environmental Stewardship Scheme • Volunteered Geographic Information • National Trails • social media

Introduction

The Environmental Stewardship Scheme (ESS) is an agri-environmental scheme (AES) in England, integrating environmental concerns into the European Commission's Common Agricultural Policy (CAP) (European Commission, 2015). The ESS provides government-financed payments to farmers and land managers in return for a commitment to farming their land with more care for the environment (Smith et al., 2013).

Introduced during 2005 and 2006, a key objective of the ESS is the maintenance and enhancement of landscape quality and character (Natural England, 2011). The ESS was established as a 'broad and shallow' approach to AES in order to extend its reach to high proportions of the countryside (Amy et al., 2013). Therefore the ESS is non-competitive and open to all farmers and land managers whose land is part of the farmed environment and is registered in the Rural Land Register (Natural England, 2013b). With agreements in place on over 70% of agricultural land in England, the ESS represents the country's most widespread approach to environmental management (Defra, 2013).

England's diverse landscapes are connected by a 4000 km network of 15 National Trails which aim to provide greater access to the English countryside (Long Distance Walkers Association, 2014), rewarding natural adventures, and the opportunity for people to be inspired by varied scenery and landscapes (Wood-Gee, 2008). In 2012 approximately 12 million visits were made to England's National Trails (Ramblers, 2012). The maintenance and enhancement of landscape quality and character is one of the primary objectives of the ESS, and evidence suggests that the ESS improves biodiversity (Defra & Natural England, 2008). Based on these observations, one could assume that the ESS can play an important role in providing a positive experience for visitors to National Trails.

Corresponding Author:

Tom Wilson[1]:

twilson@usaconservation.org, rlovelace@leeds.ac.uk, A.J.Evans@leeds.ac.uk,

[1]School of Geography, Faculty of Environment, The University of Leeds