

# Can the sentiment expressed in trail users' tweets help to assess the effectiveness of Environmental Stewardship Agreements? An exploratory analysis of the Pennine Way National Trail, England.

Tom Wilson<sup>\*1</sup> and Robin Lovelace<sup>†1</sup>

<sup>1</sup>School of Geography, University of Leeds

January 9, 2015

## Summary

This paper presents an exploratory analysis into the feasibility of using the sentiment expressed within trail users' Twitter messages to assess the effectiveness of Environmental Stewardship Scheme agreements in place along the Pennine Way National Trail, England.

**KEYWORDS:** GIS, big data, sentiment analysis, environmental stewardship, volunteered geographic information

## 1. Introduction

The Environmental Stewardship Scheme (ESS), an agri-environmental scheme, represents the most widespread approach to environmental management in England. ESS provides financial rewards to farmers and land managers in return for reductions in farming intensity and the adoption of measures to protect the surrounding environment. The success of early agri-environmental schemes was measured by levels of participation but more recently (and for the lifetime of ESS which was introduced in 2005-2006) the focus has shifted to analyse the environmental benefits provided under the scheme (Franks & Emery, 2013), for example with regard to landscape character (Natural England, 2014a), the enhancement of grassland, moorland and heath (Natural England, 2014b), bird populations (Davey et al., 2010), and the provision of ecosystem services (Defra, 2009).

England has 15 designated National Trails which pass through diverse landscapes and expanses of agricultural land, a substantial amount of which is managed under ESS. In 2012 approximately 12 million visits were made to England's National Trails (Ramblers, 2012). Despite the interactions that trail users unwittingly have with land managed under ESS there is currently no method to specifically obtain their opinions concerning the effectiveness of ESS in preserving and protecting the environment. The opinions of trail users are in fact limited to broad, large-scale qualitative surveys of visitors to the countryside in general, such as the Monitor of Engagement with the Natural Environment (MENE) which examines the adult population's engagement with the natural environment (Natural England, 2013). The National Trail User Surveys (The Countryside Agency, 2005; Natural England/Countryside Council for Wales, 2007) were discontinued in 2007.

This research aims to address this potential discontinuity of knowledge by exploring the feasibility of utilising the sentiment conveyed in trail users' Twitter messages (tweets) to assess the effectiveness of ESS specifically from the trail user perspective. This exploratory analysis will focus on the Pennine Way National Trail. We propose a methodology to geographically and lexically filter relevant tweets that originate from the proximity of the trail, and then perform sentiment analysis to extract the

---

\* gy10tlw@leeds.ac.uk

† r.lovelace@leeds.ac.uk

sentiment expressed. Finally we will determine whether this information can be used to assess the effectiveness of ESS.

## 2. Related research

Currently the effectiveness of ESS is measured with regard to the delivery of ES objectives and environmental benefits, for example in the provision of ecosystem services (Defra, 2009), the effects on landscape character (Natural England, 2014b), the ecological status of grassland, moor and heath (Natural England, 2014a), and the effects on bird populations (Davey et al., 2010). The Monitor of engagement with the Natural Environment (MENE) is a comprehensive nationwide survey which aims to measure the population's engagement with the natural environment (Natural England, 2013). The MENE survey is a representative but broad approach, and is not specific to trail users.

Sentiment Analysis (SA) is the process of extracting positive or negative opinions from text with the aim of analysing people's sentiment; this is seen as an important task, particularly since our opinions influence our behaviours, and this has led to a particular commercial interest (Liu, 2012). More recently however, the growth of the social web has led to an increased interest in SA of short informal text for the purposes of social research, to gain insights into specific events, and to study the affective dimension of the social web (Thelwall et al., 2012).

This research aims to address the potential shortfall of information regarding trails users' opinions by conducting SA of short informal text sent via the micro-blogging service Twitter. Specifically we explore whether the sentiment expressed within the tweets of trail users can be used to assess the effectiveness of ES agreements in delivering environmental benefits.

## 3. Study Area

The focus of this research is the 431 km (268 mile) long Pennine Way National Trail (PWNT), the oldest of England's 15 National Trails which officially opened on 24th April 1965 (Figure 1). Along its route the PWNT crosses agricultural land managed under ESS. For this analysis a 5km spatial buffer was drawn around the PWNT which will hereafter referred to as the PWNT corridor.

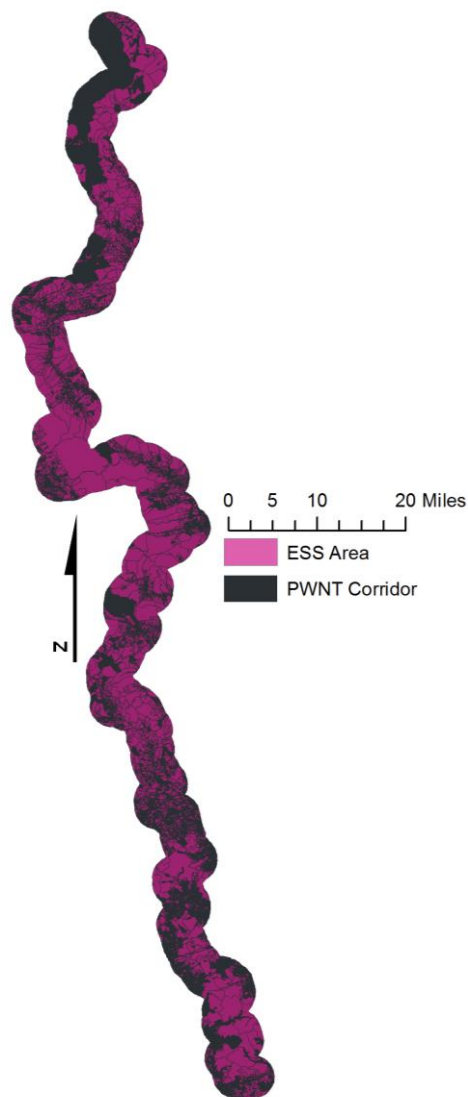


**Figure 1:** Location of the Pennine Way National Trail, England.

#### 4. Data

Geocoded Twitter data forms the basis of this research. The Twitter data were collected between 2014-06-03 and 2014-07-25 via the Twitter Streaming API, a service which provides limited access to the Twitter ‘stream’ (Lovelace et al., 2014). The full dataset consisted of 60,434 geocoded tweets which originated (i.e. were sent) from the region of the PWNT. Each row in the dataset represented a tweet and included the longitude and latitude of origin and the tweet’s text. Additional metadata included a date and time stamp, and the sender’s (tweeter’s) number of followers, following, and tweets sent. The dataset did not include a twitter username.

A GPX file of the PWNT was obtained from the National Trails website (Walk Unlimited, 2014). A 5km spatial buffer was drawn in ArcMap (ESRI, 2011) to create the aforementioned trail corridor. A shapefile of ESS agreement boundaries for England were obtained from Natural England (Natural England, 2014c). These boundaries were spatially clipped to the PWNT corridor resulting in 1717 individual ESS agreements within the trail corridor, covering 74.09% of the land (Figure 2).



**Figure 2:** The PWNT Corridor and ESS agreement boundaries. © Natural England copyright. Contains Ordnance Survey data © Crown copyright and database right [2014]

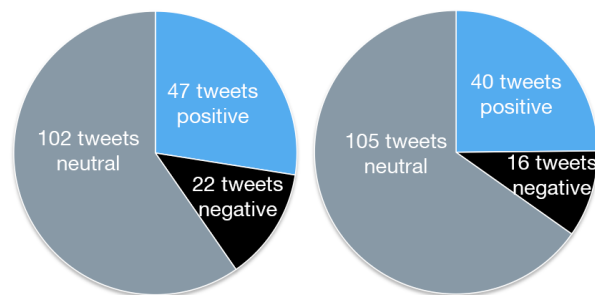
## 5. Methodology

Processing of the Twitter dataset was completed using the statistical package R version 3.1.2 (R Development Core Team, 2008). The Twitter data were spatially clipped to the PWNT trail corridor so that only tweets that originated from within the 5km PWNT corridor were included in further analysis. Regular expression terms were used to filter the tweets relevant to trail use, e.g. "Pennine Way" and "hike", plus some significant locations along the trail for example "Kinder Scout" and "Cheviots". In total 16 search terms were used. Further data 'cleaning' was performed to remove tweets such as broadcast traffic alerts and direct replies between Twitter users. This process resulted in 161 tweets for sentiment analysis.

Sentiment Analysis was conducted using SentiStrength, a lexicon-based sentiment analysis tool developed specifically for short informal text (Thelwall et al., 2010; Thelwall et al., 2012) such as that found on Twitter. SentiStrength returns an integer for both the positive (+1 to +5) and negative (-1 to -5) sentiment expressed within text. A value of 1 or -1 denotes an absence of positive or negative sentiment respectively. Overall sentiment is obtained by combining the two integers to determine sentiment polarity of the text: A combined score of 0 denotes neutral text.

## 6. Selected findings

Figure 3 provides a breakdown of the sentiment expressed in the twitter messages along the PWNT trail corridor. 47 tweets expressed positive sentiment, 22 negative sentiment and 102 expressed no sentiment (i.e. neutral). 10 tweets expressed both positive and negative sentiment. Overall tweet sentiment consisted 40 positive, 16 negative and 105 neutral.



**Figure 3:** The sentiment expressed in trail users' tweets. Left: Expressions of sentiment in each tweet. Right: The overall sentiment of the tweets

Figure 4 provides a spatial overview of the sentiment of tweets along the PWNT corridor. 7 overall positive and 3 overall negative tweets originated from within an ESS boundary.

The majority of the 161 tweets are neutral and did not express sentiment. Further investigation revealed that of the 105 overall neutral tweets 94 (90%) contain a URL to an external source such as an image or website.

## 7. Conclusions

This research has demonstrated a methodology to filter relevant tweets and perform a sentiment analysis on short informal texts sent by trail users along the Pennine Way National Trail. With regard to the feasibility of assessing the effectiveness of ESS agreements the initial findings are limited, both in terms of the number of relevant tweets and the sentiment conveyed. Additional work is needed to collect additional data, perhaps over a longer period of time. A key finding is that a high percentage of neutral tweets contained a URL to an image and it could be that these images are intended to convey sentiment. Further research should therefore focus on developing methods to extract the sentiment, if any, conveyed in these images.



**Figure 4:** Sentiment of tweets originating from within the PWNT corridor. Left: tweets expressing positive sentiment Middle: tweets expressing negative sentiment. Right: the overall sentiment of tweets. © Natural England copyright. Contains Ordnance Survey data © Crown copyright and database right [2014]

## 8. Acknowledgements

Special thanks to Dr. Andrew Evans and Dr. Paul Norman for their continued support and counsel; and to Phillapa Swanton, Hazel Thomas, and Steven Westwood of Natural England.

## 9. Biography

Tom Wilson is a GIS Online Distance Learning master's degree student at the University of Leeds, and an early-career researcher who also works as a program director of a non-profit conservation group in Arizona. His interests are in the applications of GIS, big data analysis, and geovisualisation.

Dr. Robin Lovelace is a geographer and environmental scientist, and currently a TALISMAN researcher at the University of Leeds specialising in methods of spatial microsimulation.

## 10. References

- Davey, C. M., Vickery, J. A., Boatman, N. D., Chamberlain, D. E., Parry, H. R., & Siriwardena, G. M. (2010). Assessing the impact of Entry Level Stewardship on lowland farmland birds in England. *Ibis*, 152(3), 459-474.
- Department of Food and Rural Affairs. (2009). Provision of ecosystem services through the Environmental Stewardship Scheme. Defra project code NR0121
- ESRI. (2011). ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute
- Franks, J. R., & Emery, S. B. (2013). Incentivising collaborative conservation: Lessons from existing environmental Stewardship Scheme options. *Land Use Policy*, 30(1), 847-862.
- Lawton, J.H., Brotherton, P.N.M., Brown, V.K., Elphick, C., Fitter, A.H., Forshaw, J., Haddow, R.W., Hilborne, S., Leafe, R.N., Mace, G.M., Southgate, M.P., Sutherland, W.A., Tew, T.E., Varley, J., & Wynne, G.R. (2010) Making Space for Nature: a review of England's wildlife sites and ecological network. Report to Defra.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), pp1-167.
- Lovelace, R., Malleon, N., Harland, K., & Birkin, M. (2014). Geotagged tweets to inform a spatial interaction model: a case study of museums. [online] Available at: <http://arxiv.org/abs/1403.5118>
- Natural England. (2013). Monitor of Engagement with the Natural Environment: The national survey of people and the natural environment. Annual Report from the 2011-12 survey. Natural England Commissioned Reports, Number 122
- Natural England. (2014a). Assessment of the effects of Environmental Stewardship on landscape character. Correlative analysis of datasets to assess the degree of success in delivery of Environmental Stewardship objectives. Natural England Commissioned Reports, Number 158
- Natural England. (2014b). Assessment of the effect of Environmental Stewardship on improving the ecological status of grassland, moorland and heath. Natural England Commissioned Reports, Number 156
- Natural England. (2014c). Environment Stewardship Scheme Agreements. [shapefile] Created by Natural England. August 2014. [online] Available at: <http://www.geostore.com/environment-agency/WebStore?xml=environment-agency/xml/ogcDataDownload.xml>
- Natural England/Countryside Council for Wales. (2007). Results of the national trail user survey 2007.
- Ramblers. (2012). National Trails: A fantastic future. [online] Available at: <http://www.ramblers.org.uk/what-we-do/news/2012/september/securing-a-fantastic-future-for-national-trails.aspx> [Accessed 7 January 2015].
- R Development Core Team. (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org>
- The Countryside Agency. (2005). Results of the national trail user survey 2005.
- Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *J. Am. Soc. Inf. Sci. Technol.* (63)1, pp163-173.

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), pp2544-2558.

Walk Unlimited (2014). The Pennine Way. [online] Available at: <http://nationaltrail.co.uk/pennine-way> [Accessed 7 January 2015].