How reliable is citizen-derived scientific data? An evaluation of contrail observations using multiple datasets and GIS techniques.

A. Fowler[1], J.D. Whyatt[1], R. Ellis[1], G. Davies[1]

[1]Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ
a.fowler1@lancaster.ac.uk

Summary: This paper details the preliminary outcome of research using GIS techniques and multiple datasets to evaluate the quality of contrail observations generated by citizens who participated in the Open Air Laboratories (OPAL) Climate Survey, a nationwide citizen science project. Unlike other citizen science studies in which 'experts' can revisit sites of observation, independent verification of contrail observations is difficult because of the ephemeral nature of the atmosphere. This research therefore uniquely uses datasets which reflect the complex and ever changing 3D nature of the atmosphere and have been derived from a variety of different sources (experts, models and amateur enthusiasts) with different spatial and temporal attributes.

KEYWORDS: Citizen-derived data, contrails, spatio-temporal analysis

1. Introduction

Citizen science projects encourage members of the general public to participate in scientific research (see for example: Newman et al, 2003 and Darwell and Dulvy, 1996) with citizens becoming what Goodchild (2007) recognised as 'sensors' of the world around us. They can lead to large volumes of data being collected over broad geographic areas for extended time periods (Engel and Vorshell, 2002). However, this presents the challenge of creating data of sufficient quality and accuracy to be used in scientific research. This is especially the case when volunteered geographical information does not comply with spatial data quality standards (Haklay et al, 2010). The quality of citizen-derived data can often be improved by training (MacDonald and Strachan, 1999), through financial incentives (Barreto et al, 2003) or targeted recruitment of participants (Phillips et al, 2006). However, the value of such data is often questioned by potential users.

The Open Air Laboratory (OPAL) Project is a collaboration between 15 academic and scientific institutions, led by Imperial College, London, and the Natural History Museum (http://www.opalexplorenature.org/); which enables scientists and members of the general public to contribute to scientific research around select environmental themes (see Davies et al, 2011). The OPAL Climate Survey was designed by the Met Office and Royal Meteorological Society and launched in March 2011. It comprises four selected activities designed to gain a better understanding of the interactions between human activity and the climate which cannot easily be assessed using standard scientific methods; one of these activities is observation of condensation trails or contrails. Contrails are anthropogenic cirrus clouds formed under certain atmospheric conditions when heat and water vapour emitted from aircraft exhausts mix with cool ambient air (Schumann, 2005); they can dissipate instantly or persist in the atmosphere for many hours and have been proven to contribute to regional and global temperature variations (Travis et al, 2002).

This paper introduces a method and preliminary outcomes of research to evaluate the quality of citizen-derived contrail observations. GIS techniques are used to evaluate the geographic coverage and quality of the citizen observations and verify these observations through comparison with related data sets. Unlike other citizen science studies in which 'experts' can revisit sites of observation, for example vegetation surveys and insect monitoring, independent verification of contrail observations is difficult because of the ephemeral nature of the atmosphere. This research therefore uniquely uses datasets which reflect the complex 3D nature of the atmosphere and have been derived from a variety of different sources (experts, models and amateur enthusiasts) with different spatial and temporal attributes.

1.2 OPAL Contrails Activity

Participants in the OPAL Climate Survey were required to use four reference photographs (see Figure 1) to identify the sky state. This was recorded using a standardised form allowing for an element of quality control at data entry along with the time (to the nearest 10 minutes); date (DD/MM) and postcode district (AA1). Observations could be made on any day at any time and records could be submitted by freepost, on the OPAL website or by text message. The contrails activity ran from 1st February 2011 to 30th June 2011 with 15,958 individual observations submitted in total by participants.

A: No contrails | B: Short contrails | C: Long contrails, but not spreading | D: Long-lived spreading contrails

Figure 1. OPAL contrail reference photographs

2. Methodology

The following approach was taken in line with literature suggesting a validation and then verification approach to assessing data quality (see Oreskes, 1994). An initial validation was performed to ensure all data fields required for analysis (postcode, date, time, sky state) were complete for each observation. Spatial and temporal filtering was then applied to the complete dataset to restrict observations to England for the period 05/03/2011 to 27/03/2011 and the times 09:00 to 18:00 to coincide with datasets used in verification; this resulted in 8,784 records remaining for use in analysis.

Table 1. Observations removed during filtering process

Reason for removal	Observations removed	Observations remaining
Incorrect postcode	107	15,851
Outside England	1,175	14,676
Outside date range	2,691	11,985
Outside time range	3,201	**8,784**

Given the ephemeral nature of the atmosphere, 4 different datasets were used for verification of the OPAL observations, each with different spatial and temporal properties and attributes associated with altitude. Figure 2 illustrates each of these datasets.

(a) National Contrail Observer Network

The National Contrail Network (NCON) consists of independently trained meteorological observers, who make contrail observations at 0900GMT daily using a standard observation card. There are only 5 stations currently active across England. There are differences between NCON and OPAL contrail classification schema hence NCON categories were re-coded to correspond with OPAL categories.

(b) Relative Humidity with respect to ice from the European Centre for Medium Range Weather Forecasting (ECMWF) model

Relative humidity with respect to ice (RH(ice)) was calculated from the ECMWF model outputs using Matlab at 3 hourly intervals between 09:00 and 18:00; and for those layers within the atmosphere most likely to correspond with contrail formation, between 7,500 and 15,000m. RH(ice) must be greater than 100% for contrails to persist and temperature less than -40°C; however, there is an error of approximately 10% within the model which also assumes linear regression of RH(ice) through atmosphere. There is complete coverage of England at ~28x28km resolution with the atmosphere divided into 91 vertical levels.

(c) Radarvirtuel.com

Aeroplane flight path data were downloaded at 1 minute intervals and assembled into hourly layers from the website radarvirtuel.com; this is a novel dataset, previously unused in academic research. This website provides real-time aircraft locations with altitude and call sign; however spatial coverage is variable as the website relies upon volunteers with receivers to track aircraft locations.

(d) Photos

OPAL contrail observers were encouraged to take photographs to support their observations. This is the only evidence through which to positively verify citizen observations. There is limited spatial and temporal coverage across England with only 35 photographs taken in the date and time frame for validation and supported with an accurate postcode.

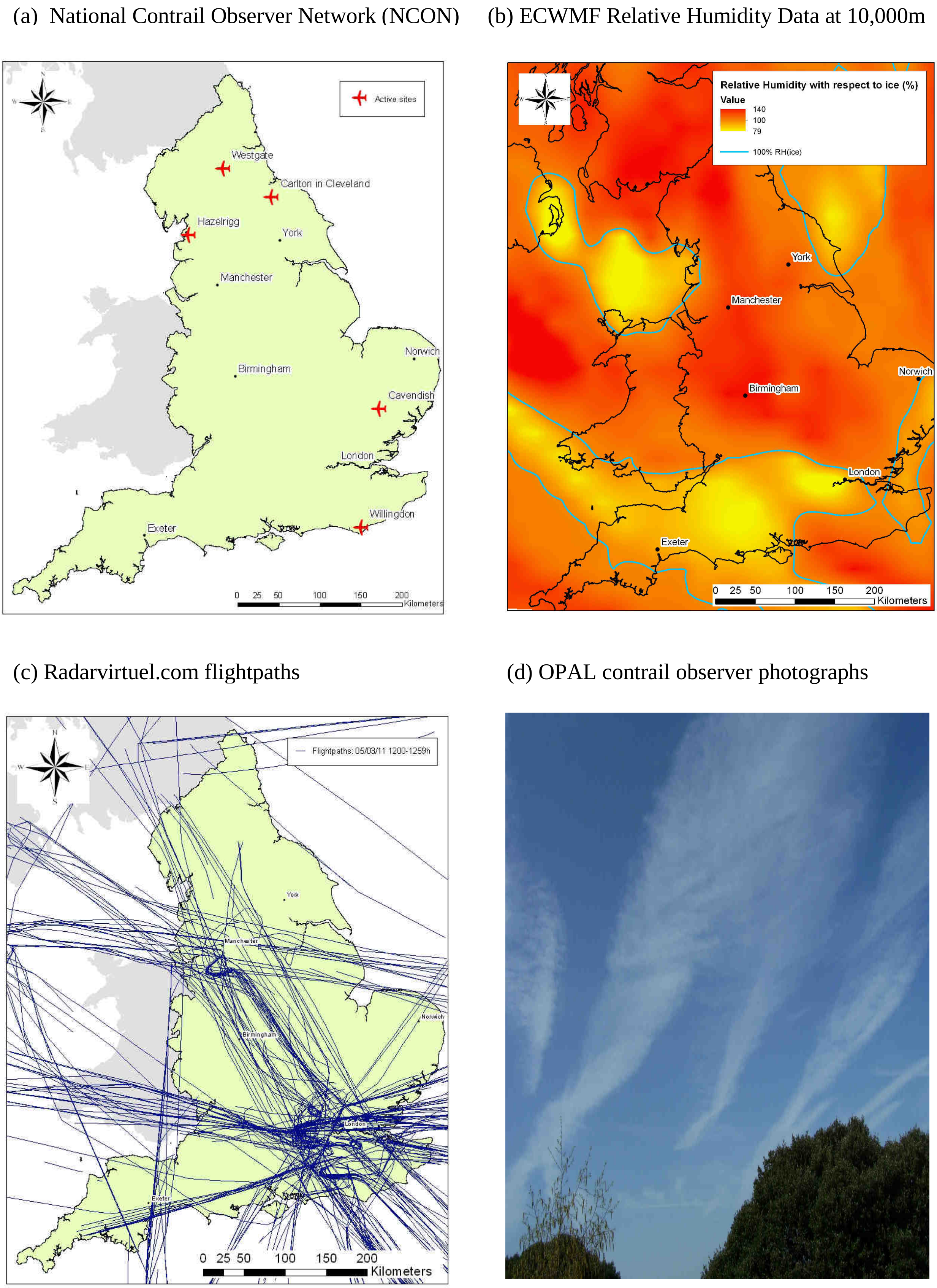
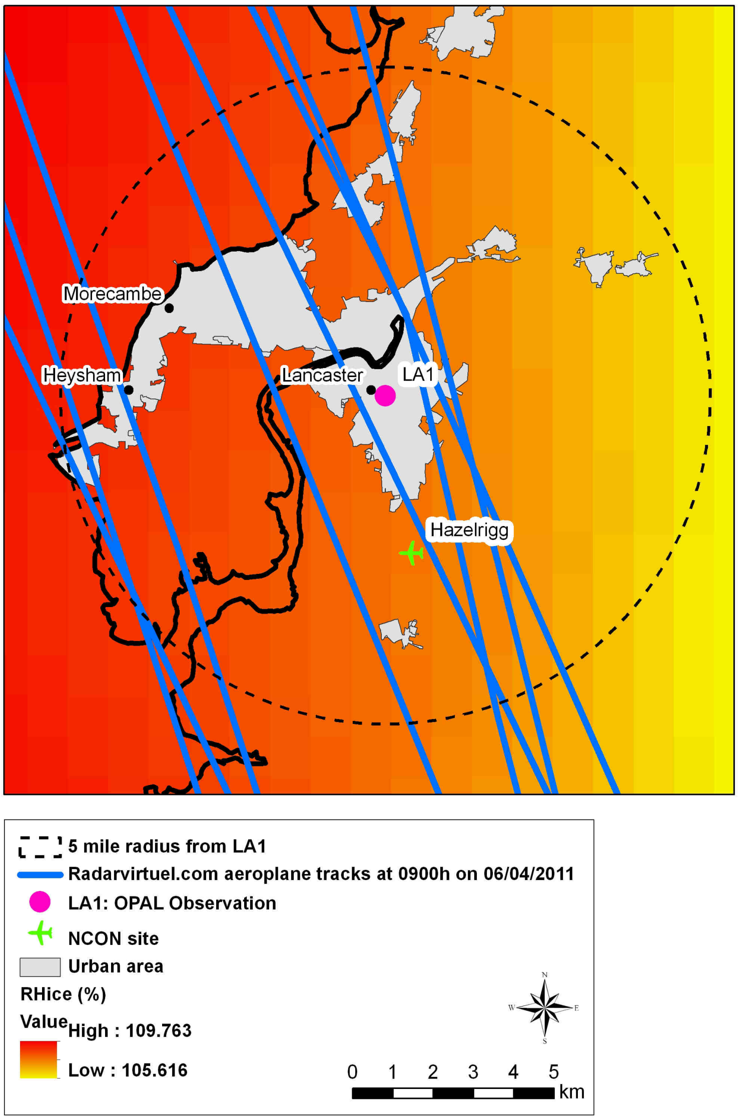


Figure 2. Datasets used in verification

3. Results

A preliminary analysis of the dataset was undertaken for one OPAL observation made within the postcode LA1 on 6th April 2011 as seen in Figure 3. The location and time of the OPAL observation define the analysis undertaken. Firstly, a buffer of 5 miles was applied to the OPAL observation at LA1 as 5 miles is considered a reasonable sky view based on observations being made in rural and urban locations. Flightpaths and NCON sites falling within this 5 mile buffer were selected and retained for analysis if their time corresponded closely with the OPAL observation (+/- 30 minutes). The relative humidity value was then extracted for the postcode centroid for levels within the atmosphere which corresponded to contrail formation.



Date: 06/04/2011
OPAL Observation: LA1 at 09:20
NCON: Hazelrigg

Altitude for contrail formation: 26,000-40,000ft

Figure 3. Example of preliminary analysis

Table 2. Outcome of preliminary analysis

Dataset	Outcome
Citizen-derived sky state	D
National Contrail Observer Network	D
Relative Humidity w.r.t ice >100% (0900-1200)	Yes (110%)
Aeroplane overhead with altitude > 26,000ft (0900-1000)	Yes
Photo	No

4. Discussion

The example presented in Figure 3 allows confidence to be placed in the citizen observation made at LA1 on 6th April 2011 at 09:20 (long lived spreading contrails confirmed). The observation of sky state D given by the OPAL participant corresponded with the NCON observation, the relative humidity value was greater than 100% and there were aircraft present above 26,000ft.

The methodology described above will be scripted using Python to verify all 8,784 observations; a metric will then be developed to assign a confidence value to each OPAL observation and statistical analysis conducted to identify bias within the citizen-derived dataset. The analysis will explore the accuracy of the OPAL dataset by comparing the observations with each of the datasets used in verification. It will then focus on comparing the OPAL observations with specific datasets; for example the variance in relative humidity per observation category.

Ultimately, the ephemeral nature of the atmosphere means greatest confidence can only be placed in OPAL observations when photographs and NCON observations are available (only for a very small % of the dataset). These two datasets provide a direct means of assessment between the OPAL participant and the trained observer. Therefore, in this analysis as a result of the lack of photographs and NCON observations, reliance will be placed on RH(ice) values and Radarvirtuel.com outputs for verification; it will be possible to generate a RH(ice) value for all OPAL observations.

The analysis of the contrails dataset will allow for conclusions to be drawn in relation to literature on citizen science and data quality and provide an indication of the level of confidence which can be placed in citizen-derived scientific data. It will also enable a judgement to be made regarding the contribution of this research to the field of GIS. The discussion will also concentrate on the limitations of ECMWF model outputs and how citizen data could inform model development and contribute to understanding the wider effects of contrail formation on regional and global climates.

The complex and ephemeral nature of the atmosphere and contrail formation makes the analysis of citizen-derived contrail observations particularly problematic. The use of GIS techniques is therefore especially important in this novel analysis as it effectively combines 3D spatio-temporal data on the atmosphere and provides additional means with which to verify citizen observations.

5. Acknowledgements

This research is funded by a NERC-ESRC Studentship.

Acknowledgement is given to the following who have assisted greatly in the research and analysis: Dr. Geoff Jenkins (Met Office & RMetS); Dr. Gaby Radel (Reading University) and Richard Forbes (ECMWF); Dr. Andy Horseman, Dr. Brian Davison, Dr. Annette Ryan, Dr. Tom Pugh (all Lancaster Environment Centre); Laurent at Radarvirtuel.com; OPAL Participants

6. References

Barreto, C., Fastovsky, D. & Sheehan, P. (2003) A model for integrating the public into scientific research. *Journal of Geoscience Education,* 50, 71-75.

Darwell, W. and Dulvy, N. (1996) An evaluation of the suitability of non-specialist volunteer researchers for coral reef fish surveys. Mafia Island, Tanzania. *Biological Conservation* 78 p223-231

Davies, L., Bell, J.N.B., Bone, J., Head, M., Hill, L., Howard, C., Hobbs, S.J., Jones, D.T., Power, S.A., Rose, N., Ryder, C., Seed, L., Stevens, G., Toumi, R., Voulvoulis, N., White, P.C.L. (2011) Open Air Laboratories (OPAL): A community-driven research programme. *Environmental Pollution* 159 2203-2210

Engel, S.R. and Voshell, J.R. (2002) Volunteer Biological Monitoring: Can It Accurately Assess the Ecological Condition of Streams? *American Entomologist* 48(3) p164-177

Goodchild, M.F. (2007) Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial Data Infrastructures Research* 2 24-32

Haklay, M., Basiouka, S., Antoniou, V., Ather, A. (2010) How many volunteers does it take to map an area well? The validity of Linus' Law to volunteered geographic information. *The Cartographic Journal* 47(4) 315-322

Macdonald, D. W. and Strachan, R. (1999) *The mink and the water vole.* Analyses for conservation. Wildlife Conservation Res. Unit and the Environment Agency, Oxford, UK.

Newman, C., Buesching, C. & MacDonald, D. (2003) Validating mammal monitoring methods and assessing the performance of volunteers in wildlife conservation- "sed quis custodiet ipsos custodies?". *Biological Conservation,* 113, 189-197.

Oreskes, N., Shrader-Frechette, K. and Belitz, K. (1994) Verification, validation and confirmation of numerical models in the earth sciences. *Science* 263 p641-646

Phillips, T., Lewenstein, B. V. & Bonney, R. (2006) A Case Study of Citizen Science. IN Cheng, D., Metcalfe, J. & Schiele, B. (Eds.) *At the Human Scale: International Practices in Science Communication.* Beijing, China, Science Press.

Schumann, U. (2005) Formation, properties and climatic effects of contrails. *C.R.Physique* 6 p549-565

Travis, D. J., Carleton, A.M. and Lauritsen, R.G. (2002) Climatology: Contrails reduce daily temperature range. *Nature* 418 p601

7. Biography

Amy Fowler is a 3rd year PhD student employing an interdisciplinary methodology to examine the interaction between the general public and scientific institutions as demonstrated by the Open Air Laboratories (OPAL) Climate Survey. Amy is funded by a NERC-ESRC Studentship and based within the Lancaster Environment Centre, Lancaster University.

Duncan Whyatt is a senior lecturer in GIS and academic supervisor to Amy Fowler. He is based within the Lancaster Environment Centre, Lancaster University.

Rebecca Ellis is a lecturer in human geography and academic supervisor to Amy Fowler. She is based within the Lancaster Environment Centre, Lancaster University.

Gemma Davies is a GIS Officer based within the Lancaster Environment Centre, Lancaster University.