Crowdsourcing for Climate and Atmospheric Sciences: *Current Status and Future Potential*

Muller, C.L.[1a], Chapman, L. [1], Johnston, S.[2], Kidd, C.[3], Illingworth, S.[4], Foody, G.[5], Overeem, A.[6], Graves, R.[7]

[1]*School of Geography, Earth & Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom*

[2]*OpenSignal, 144a Clerkenwell Road, London, EC1R 5DF*

[3]*Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20740, USA and NASA/Goddard Space Flight Center, Greenbelt, MD 20771, USA*

[4]*Manchester Metropolitan University, School of Research, Enterprise & Innovation, John Dalton Building, Chester Street, Manchester, M1 5GD,*
, United Kingdom

[5]*School of Geography, University Park, University of Nottingham, Nottingham, NG7 2RD, United Kingdom*

[6]*Hydrology and Quantitative Water Management Group, Wageningen University, Wageningen, Netherlands, and Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands*

[7]*Earth Observation Science, Physics and Astronomy, University of Leicester University Road, Leicester, Leicestershire, LE1 7RH, United Kingdom*

[a]*Email: c.l.muller@bham.ac.uk*

Abstract

Crowdsourcing is traditionally defined as obtaining data or information by enlisting the services of a (potentially large) number of people. However, due to recent innovations, this definition can now be expanded to include 'and/or from a range of public sensors, typically connected via the Internet.' A large and increasing amount of data is now being obtained from a huge variety of non-traditional sources – from smart phone sensors to amateur weather stations to canvassing members of the public.

Some disciplines (e.g. astrophysics, ecology) are already utilising crowdsourcing techniques (e.g. citizen science initiatives, web 2.0 technology, low-cost sensors), and whilst its value within the climate and atmospheric science disciplines is still relatively unexplored, it is beginning to show promise. However, important questions remain; this paper introduces and explores the wide-range of current and prospective methods to crowdsource atmospheric data, investigates the quality of such data and examines its potential applications in the context of weather, climate and society. It is clear that crowdsourcing is already a valuable tool for engaging the public, and if appropriate validation and quality control procedures are adopted and implemented, it has much potential to provide a valuable source of high temporal and spatial resolution, real-time data, especially in regions where few observations currently exist, thereby adding value to science, technology and society.

Keywords: Internet of Things, Big Data, citizen science, sensors, amateur, applications

1. Introduction

Information regarding the state of the atmosphere can now be obtained from many non-traditional sources such as citizen scientists (Wiggins and Crowston, 2011), amateur weather stations and sensors, smart devices and social-media/web 2.0. The term ‘crowdsourcing’ has recently gained much popularity; originally referring to *‘the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call’* (Howe, 2006) in order to solve a problem or complete a specific task, often involving micro-payments, or for entertainment or social recognition (Kazai *et al.*, 2013), it can now also be applied to data that is routinely collected by public sensors and transmitted via the Internet. As such, people are no longer simply consumers of data, they can also be producers (Campbell *et al.*, 2006).

These types of crowdsourcing techniques could play a vital role in the future, especially in densely populated areas, regions lacking data or countries where traditional meteorological networks are in decline (GCOS 2010). Fifty per cent of the world’s population now reside in urban areas, with this number expected to increase to 70% by 2050 (UN, 2009). Although a relatively dense network of standard *in situ* meteorological and climatological instrumentation are located in highly populated environs, cost-limitations often mean that these are not widely available in real-time or at the range of spatiotemporal scales required for numerous applications, such as: flood-water and urban drainage management (e.g. Willems *et al.*, 2012; Arnbjerg-Nielsen *et al.*, 2013), urban heat island monitoring (e.g. Tomlinson *et al.*, 2013), planning and decision-making (e.g. Neirotti *et al.*, 2014), precision farming (e.g. Goodchild, 2007), hazard warning systems (e.g. NRC, 2007), road winter maintenance (e.g. Chapman *et al.*, 2014), climate and health risk assessments (e.g. Tomlinson *et al.*, 2011), nowcasting (e.g. Ochoa-Rodriguez *et al.*, 2013), model assimilation and evaluation (e.g. Ashie and Kono, 2011), radar and satellite validation (e.g. Binau, 2012), and other societal applications. With extreme weather events expected to increase in frequency, duration and intensity in many regions in the future (IPCC, 2012), dense, high-resolution observations will be increasingly required to observe atmospheric conditions and weather phenomena occurring in more populous regions in order to mitigate future risks, as well as in less populated regions where essential data is often lacking. Indeed, Goodchild (2007, p.10) acknowledges that the most important value of such information may be in what it tells us about *"local activities in various geographic locations that go unnoticed by the world's media".*

Computing power continues to increase, doubling approximately every two years (Moore, 1965; Schaller, 1997), and with more than 8.7 billion devices connected to the internet - expected to rise to more than 50 billion by 2020 (Evans, 2011) - the amount of accessible data is growing. The 'Internet of Things' (IoT) - referring to an internet that provides "any time, any place connectivity for anything" (Ashton, 2009) - is enabling accessibility to a vast amount of data, as more devices than people are now connected to the Internet. It is predicted that the IoT could add $14.4 trillion to the global economy by the end of the decade (Bradley *et al.*, 2013), and it has great potential to improve our way of life (Gonzales, 2011). Many projects are already sourcing, mining and utilising this 'Big Data', a '*buzzword du jour'* that has become an established term over the past few years. Big Data refers to the ubiquitous, often real-time nature of data that is becoming available from a variety of sources, combined with an increasing ability to store, process and analyse such data, in order to extract information and therefore knowledge. Within the climate and atmospheric sciences - and many other scientific and mathematical disciplines - researchers are very familiar with processing and analysing large datasets, from model output to satellite datasets. However, Big Data in this sense is a term that has been created to refer to the sheer volume, velocity, variety, veracity, validity and volatility (Normandeau, 2013) of data that is now available from a range of sources. The term has been popularised and driven forward by 'smart' technologies and investment in the 'smart city' (Holland, 2008) initiative - with the term 'smart' referring to advanced, internet-enabled technology, techniques or schemes that produce informed and intelligent actions based on a range of input ('data-driven intelligence', Nielsen, 2012) - whereby populated regions are becoming equipped with various sensors (e.g. intelligent transport systems, smart (energy) grids, smart environments etc.), thereby generating a huge amount of data as well as vast scientific, operational and end-user opportunities.

With these innovations, the potential to 'source' information about a specific, localised phenomenon or variable at a high spatiotemporal resolution is at a level not previously experienced. Such data are already being used for the benefit of both the telecommunications and financial industries, with manufacturing, retail and energy applications also beginning to realise the potential that such data can provide. Crowdsourcing is already being widely used for acquiring data in other subjects (e.g. astronomy, ecology, health; Cook, 2011; Nielson, 2011), yet the realisation of the potential for utilising the data in scientific research and applications (discussed in *Section 4*) remains in its relative infancy within atmospheric science disciplines. Such data could therefore play an important role in the next age of scientific research and have numerous societal applications, but in order to determine the extent to which these non-traditional data could be incorporated, thorough quality assessments need to be conducted. Questions remain regarding the precise scientific and societal applications that could truly benefit from incorporating crowdsourced weather and climate data, how and where data should be crowdsourced from, and how the quality of this data (which is more likely to be prone to errors than those data provided by authoritative sources), can be assessed. Moreover, the issue of whether high-resolution data from smart devices and 'hidden' networks in conjunction with vast computing power, could lead to new innovations over the coming decades also needs to be addressed. Clearly crowdsourcing has the potential to overcome issues related to spatial and temporal representativeness of observations.

This paper provides an overview of crowdsourcing techniques in the context of meteorology and climatology by reviewing a number of current crowdsourcing projects and techniques, addresses uncertainties and opportunities, examines the current state of quality assurance and quality control procedures, explores future possibilities and applications, and concludes with some recommendations for these non-standard data sources that have the potential to augment and compliment existing observing systems in the future.

2. Current Approaches

Crowdsourcing traditionally relies upon a distributed network of independent participants solving a set problem. However, crowdsourcing has now moved beyond this basic approach to incorporate distributed networks of portable sensors that may be activated and maintained through the traditional protocol of crowdsourcing, such as an open call for participation, as well as repurposing data from large pre-existing sensor networks (i.e. a meteorologist deploying a network of low cost sensors specifically to examine urban climate is not crowdsourcing; whilst a meteorologist accessing data from existing amateur weather stations would be). Thus, it can be broken down into several different approaches. These can be broadly categorised as '*animate'* and '*inanimate'* crowdsourcing, with the primary distinction being the nature of the 'crowd' in question. Inanimate crowdsourcing involves obtaining or repurposing data from a range of sensors and sensor networks (e.g. sensors on streetlights, city-wide telecoms signals), whilst animate crowdsourcing requires some form of human involvement. This may result in data collection via *automated* (i.e. data is automatically collected via sensors and uploaded, though may require some form of human-intervention during installation for example), *semi-automated* (i.e. data is collected using a sensor but uploaded manually) or *manual* (i.e. human-generated data that is manually collected, entered and uploaded) means.

Alternatively, these methods could be thought of as *active* or *passive:* Active crowdsourcing (or 'human-in-the-loop sensing', Boulos *et al.*, 2011) whereby the citizen is constantly involved and is the primary processing unit that outputs data to the central node (e.g. citizen science initiatives, or utilising website, smart apps and web 2.0 platforms); Passive crowdsourcing on the other hand, is where the citizen becomes the 'gatekeeper' of their own individual sensor, installing it and ensuring its continued operation (e.g. amateur weather stations, mobile phone sensors or apps which *"silently collect, exchange and process information"* (Cuff *et al.*, 2008)). Thus, passive crowdsourcing requires no human interaction during the data collection or upload process, with citizens simply serving as regulators, whilst *semi-passive* or *semi-automated* crowdsourcing requires human-involvement if data needs to be pushed to a central server. *Figure 1* illustrates the breakdown of these different approaches, whilst *Table 1* provides an overview of some current examples of atmospheric science-related crowdsourcing approaches and projects, which are further discussed below.

2.1. Citizen Science

Citizen science is a form of collaborative research involving members of the public: volunteers, amateurs and enthusiasts (Goodchild, 2007; Wiggins and Crowston, 2011; Roy *et al.,* 2012). It can be thought of as a form of animate crowdsourcing - or 'participatory sensing' - when it actively involves citizens collecting or generating data. Hardware sensors can be used by citizens to collect data, but citizens themselves can also be classified as 'virtual sensors' by interpreting sensory data (Goodchild, 2007; Boulos *et al.*, 2011). For example, traditional eye witness reports were recently used to assess the development and movement of a series of severe thunderstorms - including hail size - across the UK on 28th July 2012 (Clark and Webb 2013).

There are many examples of citizen science projects; the Zooniverse (https://www.zooniverse.org/) and the Citizen Science Alliance (CSA; http://www.citizensciencealliance.org/) promote numerous citizen science projects, the majority of which involve data analysis rather than data creation. Some projects have been branded 'Extreme Citizen Science' since participants collect, analyse and act on information using established scientific methods (Sui *et al.*, 2013). Subjects such as ecology (e.g. NestWatch: http://nestwatch.org/; Birding 2.0: Wiersma, 2010), phenology (e.g. Natures Calendar: http://www.natuurkalender.nl/) and astronomy (e.g. Galaxy Zoo: http://www.galaxyzoo.org/) lend themselves well to such methods, with many projects finding that citizen science can generate high quality, reliable and valid scientific outcomes, insights and innovations (Trumbull *et al.*, 2000). However, its application within atmospheric science disciplines remains very much unexplored.

‘Old Weather’ (http://www.oldweather.org/) is a ‘data mining’ citizen science project aiming to help scientists recover Arctic and worldwide weather observations made by US ships since the mid-19th century by enlisting citizens to interpret old transcriptions (e.g. track ship movements) in order to generate new data. Such data can contribute to climate model projections and ultimately improve our knowledge of past environmental conditions. Similarly, the ‘Cyclone Centre’ project (http://www.cyclonecenter.org/) is utilising citizen scientists to manually classify 30 years of tropical cyclone satellite imagery.

There are also a number of citizen science programmes that actively source data directly from members of the public. For example, the GLOBE Programme (Global Learning and Observations to Benefit the Environment; http://www.globe.gov/; Finarelli, 1998) is an established, international science and education project whereby students and teachers can take scientifically valid environmental measurements and report them to a publicly available database. Since scientists can use the GLOBE data, training programmes and protocols are provided, the instrumentation involved must meet rigorous specifications and the data follows a strict quality-control procedure. Such protocols should be an imperative part of any citizen science project. In addition, the Community Collaborative Rain, Hail and Snow Network (CoCoRaHS: http://www.cocorahs.org/) is a non-profit, community-based network of volunteers who measure and map precipitation using low-cost measurement tools with an interactive website. The aim of CoCoRaHS is to provide high quality data for research, natural resource and education applications (Cifelli *et al.*, 2005). The project started in Colorado in 1998 and now has networks across the US and Canada, involving thousands of volunteers, making it the largest provider of daily precipitation observation in the US. CoCoRaHS inspired a similar project that was trialled in the UK - ‘UK Community Rain Network’ (UCRaiN) - which showed the potential for setting up a UK-based network (Illingworth *et al.*, 2014). International projects are also implementing citizen observatories for collating information about specific phenomena; for example the ‘We Sense It’ project (http://www.wesenseit.com/web/guest/home) will develop a citizen-based observatory of water to allow citizens and communities to become active stakeholders in data capturing, evaluation and communication, ultimately for flood prevention. Such networks can make real contributions to the advancement of science. For example, the National Oceanic and Atmospheric Administration’s (NOAA) ‘Precipitation Identification Near the Ground' (PING) project (Binau, 2012) is attempting to improve the dual-polarization radar hydrometeor classification algorithm, by recruiting volunteers to submit reports on the type of precipitation that is occurring in real time, via the internet or mobile phones (mPING; Elmore *et al.*, 2014), to allow radar data to be validated, whilst the European Severe Weather Database collates eye-witness reports of phenomena such as tornados, hail storms, and lightening (http://www.essl.org/cgi-bin/eswd/eswd.cgi). Furthermore, there are other forms of public crowdsourcing that go beyond measurements and observations. For example, ClimatePrediction.net is a distributed computing, climate modelling project that utilises citizen's computers to simulate the climate for the next century (http://www.climateprediction.net/).

Overall, citizen science projects are becoming an increasingly popular means to engage the public, whilst also benefiting scientific research; indeed there has been a surge in the number of citizen science projects in recent years (Gura, 2013), due to both emerging and affordable technological advances, and also the growing ubiquity of social media and new communications platforms, which offer increased accesses to participants (Silvertown 2009) as well as providing support during such projects (Roy *et al.*, 2012).

2.2. Social Media

While e-mail, SMS (Short Message Service) and web forms are the traditional means to transmit information, the recent proliferation of web 2.0 channels (e.g. the Twitter micro-blogging site, Facebook social media site, Foursquare mobile information sharing site, picture sharing sites such as Flickr and other blogs, wikis, and forums) have opened up opportunities to engage with citizens for scientific purposes, as well as for crowdsourcing data. Volunteered Geographic Information (VGI) and 'wikification of GIS' are phrases previously coined to describe the array of geo-located data that is now available from a large number of internet-enabled devices (Boulos *et al.*, 2011); social media channels are another source that can now be used to harvest an array of geo-located, date and time-stamped information (e.g. data, notes, photos, videos), which can be accessed directly (e.g. using hash-tags, key words), and in real-time.

For example, citizen-generated data has been used to monitor and map snow via social media channels. The 'UK snow map' (http://uksnowmap.com/#/) was set up to monitor and map snowfall across the UK with citizens giving the snowfall a rating out of ten which, in conjunction with a range of specific hash-tags (e.g. #UKSnowMap, #UKSnow); Muller (2013) also used social media to obtain higher-resolution snow-depths across Birmingham, UK; and in the US, the University of Waterloo's 'SnowTweets project' (http://snowcore.uwaterloo.ca/snowtweets/index.html) collates information from snow-related tweets. Storms have also been mapped using Twitter (e.g. https://ukstorm2013.crowdmap.com/), with services such as 'Twitcident' (http://twitcident.com/) monitoring, filtering and analysing twitter posts related to incidents, hazards and emergencies in order to provide real-time signals for use by police and other members of society. Mobile applications (apps) are also providing a new means to collect a range of data. Social apps are a means for citizens to submit information and there are several apps now sourcing local weather information. For example, Metwit (https://metwit.com/) is a social weather application that allows users to submit and receive information about current weather conditions using a range of weather icons (e.g. sunny, rainy, foggy, snow flurries), whilst Weddar (http://www.weddar.com/) is a 'people powered' service which asks users to indicate how they 'feel' using coloured symbols (e.g. perfect, hot, cold, freezing).

Social media can also be used in crisis management during extreme events (e.g. Goodchild and Glennon, 2010), since it enables situations to be monitored, and messages to reach key demographics quickly and efficiently. For example, one million tweets, text messages and other social media objects were used to track typhoon Haiyan and to map its damage (Butler, 2013), across the Philippines during November 2013. However, as indicated by the post-analysis of social media updates during Hurricane Irene in 2011, there is still a lot of research needed to better evaluate and inform the use and integration of social media into relief response during such extreme events (Freberg *et al.*, 2013). Furthermore, social media feeds often generate a lot of 'noise' and invalid information (Scanfeld *et al.*, 2010), which can result in biased information being amplified through the viral nature of social media misinformation (Boulos *et al.*, 2011). Therefore caution is required when utilising uncontrolled social media-generated information – both human and/or machine-based quality control, filtering and validation procedures are essential (discussed further in *Section 3*).

2.3. In situ Sensors

Whilst personal weather stations have been popular with amateur weather enthusiasts for decades, there are now an increasing number of internet-enabled, low-cost sensors and instrumentation becoming available for personal, research and operational use. Data can now be crowdsourced from dedicated sensors that are found at home, or on buildings and roadside furniture (e.g. lighting columns: Chapman *et al.* (2014); Smart Streets: http://vimeo.com/80557594) that form part of research, public or private sensor networks. These data can be transmitted via a range of communication techniques, such as Wi-Fi, Bluetooth and machine-to-machine SIM cards, contributing to the IoT and making available a large amount of data.

For example, Air Quality Egg (http://airqualityegg.com) is a community-led, air quality-sensing network that allows citizens to participate in the monitoring of nitrogen dioxide (NO_2), carbon monoxide (CO), temperature and humidity using a low-cost, internet-enabled sensor and web platform. Other low-cost sensors include Bluetooth and internet-enabled sensors - for example, infrared sensortag (Shan and Brown, 2005), rainfall disdrometers (e.g. Minda and Tsuda, 2012; Jong, 2010), air quality monitoring (e.g. Honicky *et al.*, 2008) and other sensors modified to connect to Raspberry Pi and Arduino boards (e.g. Goodwin, 2013). Numerous websites have been set up to crowdsource data from these devices – for example, tweets can be generated automatically from Air Quality Egg data, whilst websites such as Weather Underground (http://www.wunderground.com/personal-weather-station/signup), the UK Met Office 'Weather Observation Website' (WOW: http://wow.metoffice.gov.uk; Tweddle *et al.*, 2012) and the NOAA Citizen Weather Observer Program (CWOP: http://wxqa.com/) harvest amateur weather data from thousands of sites - vastly outweighing standard measurement sites - and provide hubs for the sharing and archiving of real-time and historic data (Bell *et al.*, 2013). Some of these even provide the ability to upload supplemental data ('metadata') about the location, equipment and/or data. For example, WOW uses a star rating system based on user-supplied information to indicate the quality of the data, equipment and exposure, whilst other schemes have implemented badges in recognition of expertise or data quality (Tweddle *et al.*, 2012). Furthermore, there is also freely available software (e.g. Weather Display: http://www.weather-display.com/index.php; Cumulus: http://sandaysoft.com/products/cumulus), which can display live data from a variety of low-cost sensors, as well as stream data via websites.

As a result of technological advances and the continued miniaturisation of technology, low-cost sensors are being increasingly and routinely incorporated into devices such as mobile phones, vehicles, watches and other gadgets; they are even being attached to animals (e.g. pet cameras). However, as for all forms of crowdsourcing, caution must be exercised when utilising data from such low-cost devices; analysis, calibration and inter-comparisons are required to investigate the accuracy and sensitivity of sensors rather than simply relying on the information supplied by the manufacturer.

2.4. Smart devices

Worldwide, one in every five people owns a smartphone (Heggestuen, 2013), and this figure is even higher in more economically developed countries. A large number of sensors are now being designed for connection to smart devices - for example, BlutolTemp Thermometer (EDN, 2013); iCelsius thermistor (Aginova, 2011); Plus Plugg weather sensors (http://www.plusplugg.com/en/#!); iSPEX aerosol measuring sensor (www.ispex.nl); AirCasting Air Monitor (http://aircasting.org/); Netatamo weather stations (e.g. http://www.netatmo.com/) - with projects already set up to utilise these pervasive devices. For example the N-Smarts pollution project is using sensors attached to GPS-enabled smart phones to gather data, in order to help better understand how urban air pollution impacts both individuals and communities (Honicky *et al.*, 2008).

GPS have been embedded in mobile phones for some time (since Benefon Esc in 1999) and hold much potential for applications such as distributed networks for traffic monitoring and routing (Krause *et al.*, 2008). Additional sensors are increasingly being built into these devices as standard (e.g. smart phones, tablets). For example, the Galaxy S4 contains geomagnetic positioning, as well as a gyrometer, accelerometer, barometer, thermometer, hygrometer, RGB light sensor, gesture sensor, proximity sensor and microphone (Nickinson, 2013). Data collected by these sensors can be harvested via the Internet, with this form of crowdsourcing often referred to as 'human-in-the-loop sensing' (Boulos *et al.*, 2011). For example, Overeem *et al.* (2013a) recently crowdsourced battery temperature data from mobile phones using the OpenSignal app (http://opensignal.com/). Utilising a heat transfer model, a relationship was found between daily-averaged ambient air temperatures and mobile phone battery temperatures for several cities. In addition, WeatherSignal is a smart phone app that collects live weather data by making use of the range of sensors pre-built into smart phones. PressureNet (http://pressurenet.cumulonimbus.ca/) is another app that collects atmospheric pressure measurements from its users, with the aim of using this data to help understand the atmosphere and better predict the weather. However, temperatures and other weather variables can vary significantly over small distances, especially over the heterogeneous morphology found in urban areas. This is clearly an advantage of using such sources of data, yet simultaneously highlights the potential for issues regarding data quality and reliability (e.g. errors, validations and scaling up data – discussed further in *Section 3*).

2.5. Moving platforms

Many different types of platforms are traditionally used to conduct scientific research and collect data, so the use of moving platforms is far from a new concept. What is novel is the potential for any moving platform to routinely collect information and potentially make use of existing sensors that are already built-in. The low-cost sensors mentioned above are essentially portable sensors, for example the Air Project (Costa *et al.* 2006) used citizens equipped with portable air monitoring devices to explore their neighbourhoods for pollution hotspots. Other moving platforms can also be used to collect non-fixed data. Bikes are one potential platform for crowdsourcing data (e.g. Melhuish and Pedder 2012; Brandsma and Wolters 2012). For example, Cassano (2013) used a 'weather bike' (fitted with a Kestrel 400 hand-held weather station and GPS logger) to collect temperature measurements across Colorado, finding variations of up to 10°C over a distance of 1 km, whilst the Common Scents project uses bicycle-mounted sensors to generate fine-grain air quality data to allow citizens and decision-makers to assess parameters in real-time (Boulos *et al.*, 2011). Indeed, the use of bicycles as vehicles for hosting air quality monitoring devices is becoming increasingly popular. Work by Elen *et al.* (2012) presents an air quality monitor equipped bicycle, Aeroflex, which records black carbon and particulate matter measurements as well as the geographical location. Aeroflex is also equipped with automated data transmission, pre-processing and visualisation.

Boats and ships have a long history of providing meteorological data; Since the 1940s ships have routinely collected sea surface temperature observations. Therefore all boats - commercial, military, private - provide opportunities for crowdsourcing, especially if linked to low-cost technology. For example, the International Comprehensive Ocean-Atmospheric Data Set (ICOADS) collates extensive data spanning three centuries from a range of evolving onboard observation systems, which is critical for data-sparse marine regions (Woodruff *et al.*, 1987; Worley *et al.*, 2005; Berry and Kent, 2006). Oceanographic science applications are being further explored through data obtained from low-cost, homemade conductivity, temperature and depth instruments (Cressey, 2013). A large range of atmospheric data could also be crowdsourced if other low-costs sensors were installed on ships, or by utilising data from smart devices and/or citizens on board. For example, the TeamSurv (Thornton, 2013) project is enabling mariners to contribute to the creating of better charts of coastal waters, by logging depth and position data whilst they are at sea, and uploading the data to the web for processing and display. Similarly, data can be crowdsourced from other transportation such as commercial airplanes, with further potential for emergency service helicopters, and public trains. A significant amount of data is routinely collected by aircraft, but as noted by Mass (2013) a large proportion of this potentially valuable data is currently not being used. TAMDAR (Tropospheric Airborne Meteorological Data Reporting) is collected by short-haul and commuter aircrafts, and low-level atmospheric data collected during take-off and landing could significantly benefit the forecasting of thunderstorms and other weather features, in a similar manner to AMDAR (Aircraft Meteorological DAta Relay) which is utilised for forecasting, warnings and aviation applications.

One of the most mature versions of a moving platform, in terms of crowdsourcing, research and exploration, are road vehicles. Commercial, public and personal road vehicles are beginning to contain Internet-connected sensors and have the potential to make high-resolution surface observations (Mahoney and O'Sullivan, 2013; Mahoney *et al.*, 2010), with research exploring data collected from such road vehicles already being undertaken. For example, Inrix (http://www.inrix.com/) collects data from trucks and other fleets as a source of real-time information about congestion and other issues affecting travel, whilst the Research and Innovative Technology Administration's (RITA) connected vehicle research initiative is encouraging the use of data from vehicle sensors (e.g. temperature, pressure, traction-control, wiper speed: Haberlandt and Sester, 2010; Rabiei *et al.*, 2013; Drobot *et al.*, 2010). Other studies (e.g. Aberer *et al.*, 2010; Devarakonda *et al.*, 2013; Ho *et al.*, 2009; Rada *et al.*, 2012) have used vehicles and other moving platforms to host sensors for monitoring air quality. Overall, miniaturisation of the sensors used in these studies creates opportunities for smaller mobile platforms to be used for traditional observations as well as crowdsourcing (e.g. commercial/private Unmanned Aerial Vehicles (UAVs), hot air balloons).

2.6. 'Hidden' networks

Finally, it is important to highlight the potential for repurposing data from 'hidden' networks, as a form of inanimate, passive crowdsourcing. Numerous municipal networks exist, out of sight, quietly collecting routine data for various applications (e.g. transmitting mobile phone signals, sensors on lighting columns to control light levels, city-wide traffic sensors for transport management, in-built mobile sensors for monitoring the performance of the handset). However, these have the potential to be used as proxies for monitoring other variables. For example, Overeem *et al.* (2013b) used received signal level data from microwave links in cellular communication networks to monitor precipitation in the Netherlands (Messer *et al.*, 2006; Leijnse *et al.*, 2007; Overeem *et al.*, 2013b). Other work that has used sensors for monitoring environmental variables for which they have not specifically been designed includes the use of GPS measurements from low earth orbiting satellite and ground-based instruments for monitoring atmospheric water vapour (e.g. Bentsson *et al.*, 2003; de Haan *et al.*, 2009) and Mode-S observations from air traffic control radars to observe wind and temperatures (e.g. de Haan and Stoffelen, 2012; Overeem *et al.*, 2013b). It is therefore likely that there are many other environmental uses for instruments or sensor networks that have been designed and implemented for other purposes.

3. Quality Assurance / Quality Control

Arguably the biggest challenge in incorporating crowdsourced data in the atmospheric sciences - as for other disciplines - is overcoming the barriers associated with utilising a non-traditional source of data, i.e. calibration and other quality assurance/quality control (QA/QC) issues. Clearly crowdsourcing has the potential to overcome the spatial and temporal representativeness of standard data. However, whereas the measurement quality of traditional data is not often an issue due to the use of rigorously calibrated instrumentation located in sites that adhere to strict standards, can crowdsourced data provide an acceptable level of accuracy, certainty and reliability?

Cuff *et al.* (2008) previously noted issues related to 'observer effect' and bad data processing, highlighting the need for verification when utilising public sensor data. Whilst Dickinson *et al.* (2010) stated - in reference to the ecological uses of citizen science - it *"produces large, longitudinal datasets, whose potential for error and bias is poorly understood"* and is *"best viewed as complementary".* Is this true for all crowdsourced data, or do certain types of crowdsourced data or techniques show more potential? It is likely that the utility of such data is both application and parameter-specific. In order to assess the true accuracy and value of crowdsourced data, it is clear that the quality and accuracy must therefore be assessed, particularly if is to be applied to extreme events that affect property, infrastructure and lives in the future. But how can this be achieved on a routine basis? At what spatial and temporal resolution must these studies be conducted? Is there an optimal density of 'crowdsourcing sites', after which statistical analyses and filtering can be used to extract a signal from the noise? And how much does quality vary with source or product?

The great potential of crowdsourcing as a source of data is strongly tempered by concerns with its quality. The latter arises mainly because the data are typically not acquired following ‘best practices’ in accordance to authoritative standards, and may come from a variety of sources of variable and unknown quality. In the absence of information on the quality of crowdsourced data it may be tempting to use inputs from a large number of contributors, as a positive relationship between the accuracy of contributed data and number of contributors has been noted in the literature (e.g. Raymond, 2001; Flanagin and Metzger, 2008; Snow *et al.*, 2008; Welinder *et al.*, 2010; Girres and Touya, 2010; Haklay *et al.*, 2010; Heipke, 2010; Goodchild and Glennon, 2010; Goodchild and Li, 2012; Basiouka and Potsiou, 2012; Neis *et al.*, 2012; Comber *et al.*, 2013; Foody *et al.*, 2013; See *et al.*, 2013). This may not, however, always be appropriate as the accurate contributions may be lost within a large volume of low quality contributions. Indeed, there is some evidence that indicates that it can be unhelpful to have too many contributors, with accuracy declining as more data are made available (Foody *et al.*, 2014). This issue has some similarity to the curse of dimensionality which is widely encountered in satellite remote sensing, which often leads to a desire to reduce the size of the data sets in order to achieve high accuracy (Pal and Foody, 2010). The ability to rate sources of data may allow a focus on the higher quality contributions that result in the production of more accurate information (Foody *et al.*, 2014).

A variety of methods have been applied to assess the accuracy of crowdsourced data (Raykar and Yu, 2011, 2012; Foody *et al.*, 2014). In relation to crowdsourced data on geographical phenomena, a range of approaches to quality assurance are possible (Goodchild and Li, 2012). For example, the contributions from highly trusted sources or selected gatekeepers might be used to support quality assurance. Furthermore the geographical context associated with contributions may be used to check the reasonableness of the data provided by a source given existing knowledge (Goodchild and Li, 2012). There is also considerable interest in intrinsic measures of data quality that indicate features such as its accuracy, which can be obtained from the data set itself (Hacklay *et al.*, 2010; Foody *et al.*, 2014). These approaches can, in certain circumstances, allow the accuracy of the individual data sources to be assessed (Foody *et al.*, 2013, 2014). They have, however, typically been based on categorical data, therefore research into methods more suited to higher level, more quantitative data, such as that used in characterising atmospheric properties, would be required.

For temperature studies, such as detailed investigation of the Urban Heat Island (UHI) effect, it is important to have a good spatiotemporal coverage, but it is also imperative that the data is accurate and representative. For example, existing, in-built car thermometers have the potential to provide high spatiotemporal resolution data, however the accuracy of this data is questionable since quality will vary between vehicles (e.g. variety of car makes, models, and ages; different sensors of varying precision and quality, located in different parts of the vehicle; varying microscale morphological information). However, by using smart technologies and standardising instrumentation, the utility of such data appear to show potential. For example, the NCAR (National Centre for Atmospheric Research) Vehicle Data Translator (VDT) has started to extract and process data from vehicular sensors with the long-term aim to obtain data from millions of connected vehicles in an operational setting. The VDT is a modular framework designed to ingest observations from vehicles, combine it with ancillary data, conduct quality checks, flag data, compute statistics and assess weather conditions (Drobot *et al.*, 2009; 2010). Anderson *et al.* (2012) recently tested air temperature measurements from 9 vehicles (two vehicle models) over a 2-month period, these data were then run through the VDT and a 2 °C difference between the vehicle data and the measurement from the nearest (<50 km radius) ASOS (Automated Surface Observing System) station reading was used to flag suspect data, the outcome of which was that a consistent agreement with weather stations was found at this relatively coarse spatial scale. This also highlights the issue of scale and the importance of understanding what data is actually being crowdsourced (e.g. microclimate vs. local-scale vs. mesoscale; Oke, 2004; Muller et al., 2013a) in order to utilise data for appropriate applications.

Furthermore, as mentioned, smart phones have also been used to indirectly estimate temperature data at high-resolutions. However, the relationships Overeem *et al.* (2013a) found between ambient air temperatures and smart phone battery temperatures were averaged across entire cities and over whole days, therefore the utility of smart phones for higher resolution UHI analysis, for example, is still to be explored. Indeed, initial analyses in Birmingham, UK, indicated that using more appropriate representative local data for validating crowdsourced data shows promise since the accuracy of mobile temperature data that were validated using local urban weather stations showed improvement over readings validated using data from a more remote, less representative climate station (*figure 2*). However, this may also be due to using higher-precision data for the validation. Therefore, in order to fully explore this, a larger number of participants are needed to supply data before higher-resolution (in both time and space) investigations can be conducted using a high-resolution urban meteorological testbed for validation (Chapman *et al.*, 2012).

For parameters such as precipitation - which can vary significantly over short distances (e.g. 30-40% over 1-2 miles: Doesken and Weaver, 2000) particularly for convective rainfall - extra information gained from crowdsourcing could indeed provide essential data to supplement global *in situ* rainfall networks (*figure 3*), many of which are on the decline (Walsh, 2012; Lorenze and Kunstmann, 2012; Yatagai *et al.*, 2012; Tahmo, 2013; Kidd *et al.*, 2014). For example, in the US the CoCoRaHS and PING programmes provide high quality data used for research, natural resource and education applications (Cifelli *et al.*, 2005); indeed data from PING are already being used to improve the dual-polarization radar hydrometeor classification algorithm. Moreover, there is potential for more unusual-yet-pervasive platforms to be utilised for monitoring rainfall; umbrellas with built-in piezo sensors that measure raindrop vibrations on the canvas and transmit data to smart phones via Bluetooth - or 'smart brollies' - are being explored for crowdsourcing rainfall data at ground-level (Hut *et al.*, 2014).

Wind can also vary significantly over short distances, particularly in areas with high roughness length (e.g. street canyons, forests) and crowdsourcing may prove useful. However, as was found to be the case for amateur weather stations, in order for data to be reliable, details about the site of the instrumentation need to be known (Steeneveld *et al.*, 2011; Wolters and Brandsma, 2012; Bell *et al.*, 2013), although Agüera-Pérez *et al.* (2014) did find that useful wind descriptions could be generated using high-density stations - run by various public institutions - based on quantity rather than quality. Other variables may only benefit significantly from supplementary crowdsourced data for certain applications; for example pressure does not tend to vary significantly over short distances except during the passage of a front or convective bands. Madaus *et al.* (2014) recently found that assimilating additional pressure tendency data from privately owned weather stations reduced forecast error for mesoscale phenomena, offering potential for other crowdsourced data such as dense barometric readings from smart phones for the real-time tracking of storms. Therefore extreme weather phenomena that exhibit significant pressure and wind variations (e.g. tornados, hurricanes) could perhaps benefit from other forms of crowdsourced data, but at present it is difficult to determine which particular technique would be most suitable for observing such an extreme event.

Concentration of atmospheric pollutant species can also vary significantly. Very low-cost air quality sensors, such as the Air Quality Egg, iSPEX aerosol measuring sensor and AirCasting Air Monitor, are becoming more popular with members of the public. However, due to their low-cost nature, trade-off between quality and quantity is often necessary. For example, Air Quality Egg does not calibrate all of the sensors prior to shipping; instead they rely on making use of the potentially large network of sensors to compensate for a large range of readings from individual sensors (AirQualiyEgg, 2014). However, the problem with this is that it is difficult to determine whether the sensors are measuring extreme values (due to its location next to a pollutant source, for example) or whether there is a problem with the sensor.

Evidently, methods for assessing crowdsourced data are beginning to emerge (e.g. Honicky *et al.* (2008) discussed a Gaussian, process-based noise model for handling non-uniform sampling and imprecision in mobile sensing) but there are also many techniques and lessons that can be learned from other fields and disciplines. For example, satellite validation techniques, model performance evaluation methods, calibration techniques for *in situ* instrumentation (e.g. Young *et al.*, 2014). Furthermore, different crowdsourcing techniques each have their own issues, for example human error or bias, low-cost instrumentation precision and accuracy, amount of data/coverage/spatial heterogeneity (bias towards populous areas), differing amount of metadata that can be provided, varying level of data-processing, network issues (e.g. stability, availability, time-delay), varying data types and descriptions, and privacy. Metadata is therefore important for interpreting data. It is already collected for standard meteorological stations and UMNs (e.g. Muller *et al.*, 2013a; 2013b) and it is logical that metadata would also accompany crowdsourced data. However, standards and protocols for this do not currently exists; at most it is simply geographic and timestamp information that is provided with data, whereas for atmospheric variables and applications, information (e.g. local and microscale conditions, sensor details etc.) are useful or even essential for evaluation purposes. Some amateur observations website have started to encourage contributors to supply detailed supplementary information (e.g. UKMO WOW; Meteoclimatic: http://www.meteoclimatic.com/), however it is not usually obligatory to supply complete metadata. Metadata is especially important for moving sensors, and location sensing is a developing technology. The potential for sensor combination is evolving, e.g. by allowing the mobile phone itself to identify its context through the use of multiple sensors. For example, Google have a new API called 'Activity Recognition' that recognises whether the user is walking, cycling or in a vehicle, using the movement pattern recorded by the accelerometer and other sensors (Robinson, 2013). Other applications include using light sensors on mobiles to determine outdoor readings (Johnston, 2013), and the use of barometer readings to determine change in height. Thus, sensors or devices could simultaneously collect data *and* metadata, allowing for more effective cleaning of the dataset. To this end, timestamps and geo-location data are crucial.

4. Applications and Potential Innovations

If indeed the accuracy of a range of crowdsourced data can be assessed for different types, scales and quantities of data, and if protocols are put in place to monitor data quality and ensure that all the relevant supplementary information is supplied, what, therefore, is the value and utility of crowdsourced data? As discussed earlier, there are a number of applications that may indeed benefit from the increased spatiotemporal resolution and real-time nature of measurements made available by these forms of data-sourcing techniques; whereas other applications may find the quality and reliability of the data to be too poor and/or may not provide any further benefit to the standard techniques that are already utilised. An overview of some of the potential applications of crowdsourced data are outlined in *table 2*.

Weather forecasting models have already been developed to utilise a range of crowdsourced data in an attempt to provide highly localised, minute-by-minute forecasts ('nowcasts'). For example, the IBM 'Deep Thunder' micro forecasting technology (http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepthunder/) is a targeted weather forecasting program which uses a range of public weather data from NOAA, NASA, the U.S. Geological

Survey, WeatherBug and other weather sensors. Other similar apps include SkyMotion (http://skymotion.com), Dark Sky (http://darkskyapp.com/), RainAware (http://www.rainaware.com/), Nooly (http://www.nooly.com/) and TruPoint (http://www.weather.com/encyclopedia/trupoint.html). However, the accuracy of models and other products utilising amateur, crowdsourced data are very much reliant on the quality of the observations, reemphasising the need for quality control. There are many potential societal, environmental and economic applications of crowdsourced data (*table 2*) - including public health (e.g. OpenSense air quality monitoring: Aberer *et al.*, 2010), infrastructure (e.g. Climate resilience: Chapman *et al.*, 2013), education (e.g. DISTANCE IoT project: www.iotschool.org; Pham, 2014), transportation (e.g. Ad hoc networks for urban routes: Ho *et al.* 2009), winter road management and flood management (e.g. Smart Streets project: www.smartstreethub.com; Chapman *et al.*, 2014); energy (e.g. Farhangi, 2010; Agüera-Pérez *et al.*, 2014); other societal uses (e.g. Urban Atmospheres: http://www.urban-atmospheres.net) – and therefore real opportunities for utilising it to improve our way of life. Indeed, with continuous technological advances, miniaturisation of sensors, improvements to hardware and software involved in data transmission, processing and storage, and availability of 'free' internet connections (Muller *et al.*, 2013a), infrastructure and devices are becoming even smarter, which will result in a multitude of future possibilities. For example, the possibility of crowdsourcing weather using Google glass (Sheehy, 2013) or webcams; the potential to utilise data from sensors built into smart lighting columns (e.g. LUX sensors on modern lampposts) or even the use of Wi-Fi within city-wide infrastructure to upload data (e.g. the use of Smart bus-stops); routine upload of data from cars (e.g. windscreen wipers, brake pads etc) and smart phones.

Furthermore, there will be scope for utilising other forms of platforms in the future. For example, Unmanned Aerial Vehicles (UAVs), once the preserve of targeted meteorological research, are another platform that may be increasingly used since they show potential for various applications such as CCTV, filming sporting events, delivery vehicles (e.g. 'Prime Air': Amazon, 2013). They are becoming increasingly sophisticated and miniaturised, with much potential for hosting a range of sensors. If they are used more routinely in the future, these platforms and others (e.g. hot air balloons: de Bruijn, 2013) hold further potential for crowdsourcing data (e.g. for use in real-time monitoring, management, planning) in a similar way to vehicles and other moving platforms.

5. Conclusions and Recommendations

Some traditional meteorological networks are in decline (GCOS 2010), yet the demand for real-time, high spatiotemporal resolution data is increasing; therefore there is a clear need for crowdsourcing weather and climate data. Non-traditional data are now being harvested from a large number of sources at high resolutions, and the amount of crowdsourced data is only going to increase with time.

As computing power increases, our ability to process and utilise this Big Data will also increase, therefore we must explore its potential. Whilst some fields (e.g. land mapping) have already shown evidence of the value of crowdsourcing, for the atmospheric science community, in the near future at least, it will rarely be a replacement for traditional sources of atmospheric data. It could, however, become a useful, cost-effective tool for obtaining supplemental, higher-resolution information for a range of applications, especially in economically developing countries or areas containing few weather stations. In order to determine the precise benefit of utilising such data as well as the amount of validation needed, a thorough analysis of the spatiotemporal scales required and the acceptable precision and accuracy for a range of parameters, applications and/or geographic regions is required. For example, what are the spatial and temporal scales and errors required for monitoring the UHI compared to pluvial flash flooding? Five-minute resolution data may be required for urban hydrological applications, whilst hourly data may be acceptable for other regional hydrological applications. Similarly, the density of air temperatures measurements needed for observing the UHI will vary according to the urban morphology of a city (Stewart and Oke, 2013). A comprehensive assessment of this is beyond the scope of this paper, but would be extremely useful for future crowdsourcing endeavours.

However, in order for progress to be made, thorough verification and quality-checking procedures must be in place. To-date only a few studies have begun exploring the accuracy and quality of crowdsourced atmospheric data, and even fewer at high spatiotemporal resolutions. In order to validate such crowdsourced data at a high spatiotemporal scale, standardised, calibrated and quality-checked, high resolution UMNs and air quality networks are required. Such test beds may only be required in a small number of regions in order to verify crowdsourced data prior to use elsewhere. Others have also highlighted this need; for example, Boulos *et al.* (2011) stated that eradicating or lessening the issues related to crowdsourced data can be achieved by the verification of data with other sensor nodes, but acknowledged that this would depend on the density of network and the existence of other related data, which in turn depends on the requirements for each parameter or application. In a recent study, Young *et al.* (2014) installed a network of low-cost air temperature sensors within an urban weather station test bed in Birmingham, UK (Chapman *et al.*, 2012). This test bed was designed for UHI analysis, so is ideal for assessing the ability of this sensor for UHI monitoring.

Furthermore, in order to achieve a high-level of reliability, specific guidelines, standards and protocols are required to enable interoperability and in order to quantify the reliability of crowdsourced data (e.g. metadata protocols: Muller *et al.*, 2013b; QA/QC procedures: Boulos *et al.*, 2011). Current crowdsourcing projects could act as catalysts for such an international movement and encourages the use of such data by a range of end-users. Indeed, national meteorological services could even collect, verify and distribute crowdsourced data (and metadata) from separate projects and eventually integrate data via a co-ordinated initiative in order to encourage open data sharing and standardisation. Such schemes may indeed set the foundation for a future 'data web' (Nielsen, 2012).

It is also important to acknowledge the ethical implications of crowdsourcing, which depend heavily on the type of crowdsourcing in action, and the extent to which the data could be used to individually identify either the contributor or individuals exposed to the sensor network. In participatory crowdsourcing there is often a distinct contract between the individual and the organisers therefore many of the usual concerns about data collection, storage and dissemination do not apply since there is specific consent by the user to provide data to a central location for processing. However, there are a few issues related to user privacy, primarily the ability to identify people by very few location points (Montjoye *et al.*, 2012). It is therefore necessary to keep raw data private, and only publish data that does not show which device is contributing (and perhaps apply some small degree of distortion to location, whilst keeping information such as device type). Nevertheless, since crowdsourcing from members of the public is such a specific transaction that relies on participation and comprehension, it means that most privacy concerns are reduced to basic data security – provided that the organisers make clear the type of data that is being collected and its intended purpose or future use, as well as making a commitment to only making publicly available non-identifying data. A full examination of this is beyond the scope of this paper, but readers are referred to Nissenbaum (2004) for a discussion about how expectation of privacy is dependent upon the transactional context, including the ways in which it is disseminated post-transaction.

Public engagement is also a positive side effect of many types of crowdsourcing. Indeed, the contribution to science and society as well as the appreciation, wonder and connection to the natural world are key motivations for many people to become involved in such projects (Roy *et al.*, 2012). However, some schemes further incentivise people by using rewards (e.g. monetary payment), or by using 'gamification' devices such as league tables to appeal to the competitiveness of participants (Hochachka *et al.*, 2012) [1]. Therefore, at the very least crowdsourcing is a tool to engage the general public; at most it is an important source of valuable, real-time, high-resolution information where none previously existed.

Nevertheless, with improving technology and connectivity, the miniaturisation of devises and lower-costs, the 'Internet of Everything' is inevitable; We need to determine how we can take advantage of this source of data for a variety of applications such as scientific research, education, policy

[1] It is worth noting, however, that the different motivations of contributors can impact on accuracy; for example, there is some evidence that those motivated by money are more accurate - if the amount is sufficient - than those who contribute out of enjoyment (Kazai *et al.*, 2013).

generation, environmental monitoring, and societal applications. Crowdsourcing as a research field has great potential to bridge the gap between the social scientists, computer scientists and physical and environmental scientists, thereby encouraging interdisciplinary working and enhancing knowledge exchange and scientific discovery (Wechsler, 2014). However, due to the immature nature of this source of data, this review has inevitably raised more questions than answers. It is expected that over the coming years, the field will move on considerably and more of these queries will be resolved in due course. Is this truly the start of a new and valuable age of 'society in science', or is crowdsourcing simply an *en vogue* technique? For atmospheric science disciplines, time will tell whether or not it is just a lot of 'hot air'**.**

Acknowledgements

The authors would like to thank the UK Natural Environmental Research Council (NERC) for funding Dr Catherine Muller (NE/I006915/1), Dr Lee Chapman (NE/I006915/1), and Dr Illingworth (NE/I029293/1); The Netherlands Technology Foundation STW for supporting Dr Aart Overeem (project 11944); The UK NERC National Centre for Earth Observation (NCEO) for supporting Dr Rosemarie Graves, who also holds an Open Knowledge Foundation Panton Fellowship; Samuel Johnston would like to acknowledge James Robinson, CTO OpenSignal. The Authors would also like to thank Dr Andrew Hudson-Smith at the Centre for Advanced Spatial Analysis, University College London, and Dr Geoff Jenkins from the UK Met Office Hadley Centre for Climate Prediction and Research for their comments and insights whilst preparing this manuscript.

References

Aberer K, Sathe S, Chakraborty D, Martinoli A, Barrenetxea G, Faltings B, and Thiele L. 2010. OpenSense: open community driven sensing of environment. In *Proceedings of the ACM SIGSPATIAL International Workshop on GeoStreaming (IWGS '10)*, New York, NY, USA

Agüera-Pérez A, Palomares-Salas JC, de la Rosa JJG, Sierra-Fernandez JM. 2014. Regional wind monitoring system based on multiple sensor networks: A crowdsourcing preliminary test, *Journal of Wind engineering and Industrial aerodynamics*, **127:** 51-58

Ainova (2011) iCelsius User Guide [www] http://www.icelsius.com/docs/iCelsius%20BBQ%20-%20User%20Guide%20-%20rev.%202.pdf (accessed December 2013)

AirQualityEgg (2014) Data Quality [www] http://airqualityegg.wikispaces.com/Data+Quality (Accessed December 2013)

Amazon (2013) Amazon Prime Air [www] http://www.amazon.com/b?node=8037720011 (Accessed

January 2014)

Anderson ARS, Chapman M, Drobot SD, Tadesse A, Lambi B, Wiener G, Pisano P. 2012. Quality of Mobile Air Temperature and Atmospheric Pressure Observations from the 2010 Development Test Environment Experiment, *Journal of Applied Meteorology and Climatology*, **51**: 691-701

Arnbjerg-Nielsen K, Willems P, Olsson J, Beecham S, Pathirana A, Bülow Gregersen I, Madsen H and Nguyen V-T-V. 2013. Impacts of climate change on rainfall extremes and urban drainage systems: a review, *Water Science and Technology*, **68**(1): 16–28

Ashie Y and Kono T. 2011. Urban-scale CFD analysis in support of a climate-sensitive design for the Tokyo Bay area. *International Journal of Climatology*. **31:** 174–188

Ashton K. 2009. That 'Internet of Things' Thing, in the real world things matter more than ideas. RFID Journal [www] http://www.rfidjournal.com/articles/view?4986 (Accessed October 2013)

Basiouka S and Potsiou C. 2012. VGI in Cadastre: a Greek experiment to investigate the potential of crowd sourcing techniques in Cadastral Mapping, *Survey Review*, **44** (325): 153-161

Bell S, Cornford D, Bastin L. 2013. The state of automated amateur weather observations. *Weather*. **68**(2): 36-41.

Bengtsson L, Robinson G, Anthes r, Aonashi K, Dodson A, Elgered G, Gendt G, Gurney R, Jietai M, Mitchell C, Mlaki M, Rhodin A, Silvestrin P, Ware R, Watson R, Werger W. 2003. The use of GPS measurements for water vapour determination, *Bull. Am. Meteorol. Soc.* **84**(9): 1249-1258

Berry DI and Kent EC. 2006. A new air-sea interaction gridded dataset from ICOADS with uncertainty estimates, *Bull. Am. Meteorol. Soc.* **90**, 645–656.

Binau S. 2012. The PING Project [www] http://www.erh.noaa.gov/iln/ping.php (accessed 14th September 2013) http://www.nssl.noaa.gov/projects/ping/

Boulos MNK, Resch B, Crowley DN, Breslin JG, Sohn G, Burtner R, Pike WA, Jezierski E, Chuang K-YS. 2011. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples, *International Journal of Health Geographics*, **10**: 67

Bradley J, Barbier J, Handler D. 2013. Cisco White Paper: Embracing the Internet of Everything To Capture Your Share of $14.4 Trillion [www] http://www.cisco.com/web/about/ac79/docs/innov/IoE_Economy.pdf (Accessed December 2013)

Brandsma T, Wolters D. 2012. Measurement and Statistical Modeling of the Urban Heat Island of the City of Utrecht (the Netherlands). *J. Appl. Meteor. Climatol.*, 51, 1046–1060.

Butler D. 2013. Crowdsourcing goes mainstream in typhoon response, *Nature*,

doi:10.1038/nature.2013.14186
Campbell AT, Eisenman SB, Lane ND, Miluzzo E, Peterson RA. 2006. People-centric urban sensing, *Proceedings of the 2nd annual international workshop on Wireless internet*, WICON '06 New York, NY, USA
Cassano JJ. 2013. Weather Bike: A bicycle based weather station for observing local temperature variations, *Bulletin of the American Meteorological Society*, doi:10.1175/BAMS-D-13-00044.1
Chapman L, Azevedo JA. and Prieto-Lopez T. 2013. Urban heat and critical infrastructure networks: a viewpoint. *Urban Climate* **3**:7-12
Chapman L, Muller CL, Young DT, Cai X-M, Grimmond C.S.B. 2012. An introduction to the Birmingham urban climate laboratory. In: *Proceedings 8th International Conference on Urban Climates*, Dublin, Ireland. 6 - 10 August.
Chapman L, Muller CL, Young DT, Rose P, Lucas C, Walden J. 2014. Winter Road Maintenance and the Internet of Things, *17th International Road Weather Conference*, Andorra, 30th January - 1st February 2014
Cifelli R, Doesken N, Kennedy P, Carey LD, Rutledge, SA, Gimmestad C, Depue T. 2005. The community collaborative rain, hail, and snow network: Informal education for scientists and citizens. *Bulletin of the American Meteorological Society*, **86**(8): 1069-1077.
Clark MR and Webb JDC. 2013. A severe hailstorm across the English Midlands on 28 June 2012, *Weather*, **68** (11): 284–291
Comber A, See L, Fritz S, Van der Velde M., Perger C, Foody G. 2013. Using control data to determine the reliability of volunteered geographic information about land cover, *International Journal of Applied Earth Observation and Geoinformation*, **23**, 37-48
Cook G. 2011. How Crowdsourcing is Changing Science [www] http://garethcook.net/how-crowdsourcing-is-changing-science/ (Accessed March 2014)
Costa B, Schulte J and Singer B. 2006. Air Project [www] http://www.pm-air.net/ (accessed Jan 2014)
Cressey. D. 2013. Crowdsourcing may open up ocean science, *Nature*, doi:10.1038/nature.2013.13341
Cuff D, Hansen M, Kang J. 2008. Urban sensing: out of the woods, *Communications of the ACM*, **51**(3), 24-33
de Bruijn, EIF. 2013. Wind information derived from hot air balloon flights for use in short term wind forecasts [www] http://www.knmi.nl/publications/fulltexts/de_bruijn_de_haan__knmi__ems13.pdf (Accessed March 2014)
de Haan S, Holleman I, Hotslag AAM. 2009. Real-time water vapour maps from a GPS surface

network: Construction, validation and applications, *J. Appl. Meteorol. Climatol.*, **48** (7): 1302-1316

de Haan S, Stoffelen A. 2012. Assimilation of high-resolution Mods-S wind and temperature observations in a regional NWP model for nowcasting applications, *Weather Forecast*. **27** (4): 918-937

de Jong S. 2010. Low cost disdrometer: Improved design and testing in an urban environment, Masters Thesis, TU Delft [www] http://www.citg.tudelft.nl/fileadmin/Faculteit/CiTG/Over_de_faculteit/Afdelingen/Afdeling_watermanagement/Secties/waterhuishouding/Leerstoelen/Waterbeheer/onderzoek/Projects/Msc_Research/Completed/000_Jong,_S.A.P._de/doc/101207Disdrometer.pdf (accessed Dec 2013)

Devarakonda S, Sevusu P, Liu H, Liu R, Iftode L, Nath B. 2013. Real-time Air Quality Monitoring Through Mobile Sensing in Metropolitan Areas, *UrbComp'13*, August 11–14, 2013, Chicago, Illinois, USA.

Dickinson JL, Zuckerber B, Bonter, DN. 2010. Citizen Science as an Ecological Research Tool: Challenges and Benefits, *Annual Review of Ecology, Evolution, and Systematics*, **41**:149–72

Doesken NJ and Weaver JF. 2000. Microscale rainfall variations as measured by a volunteer network, *12th Conference on Applied Meteorology*, 8-12 May 2000, Asheville, NC,

Drobot SD, Chapman MC, Schuler E, Wiener G, Mahoney III WP, Pisano PA, and McKeever B. 2010. Improving road weather hazard products with vehicle probe data—The Vehicle Data Translator quality-checking procedures. *Transportation Research Record 2169, Maintenance Services and Surface Weather*, 128–140.

Drobot SD, Mahoney WP, Pisano PA, and McKeever BB. 2009. Tomorrow's forecast: Informed drivers. *ITS Int.*, 15, NA1–NA2.

EDN. 2013. Bluetooth temperature sensor works with smart phones [www] http://www.edn.com/electronics-products/other/4412268/Bluetooth-temperature-sensor-works-with-smart-phones (accessed December 2013)

Elen B, Peters J, Van Poppel M, Bleux N, Theunis J, Reggente M, and Standaert A. 2012. The Aeroflex: A Bicycle for Mobile Air Quality Measurements, *Sensors*, **13**: 221-240

Elmore KL, Flamig ZL, Lakshmanan V, Kaney BT, Farmer V, Reeves HD, Rothfusz LP. 2014. mPING: Crowd-Sourcing Weather Reports for Research, *Bulletin of the American Meteorological Society*, doi: http://dx.doi.org/10.1175/BAMS-D-13-00014.1

Evans D. 2011. The Internet of Things: How the Next Evolution of the Internet Is Changing Everything [www] http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf (Accessed November 2013)

Farhangi H. 2010. The path of the smart grid, *IEEE power & energy magazine*, **8**(1): 18-28

Finarelli MG. 1998. GLOBE: A Worldwide Environmental Science and Education Partnership, *Journal of Science Education and Technology*, 7(1)

Flanagin AJ and Metzger MJ. 2008. The credibility of volunteered geographic information, *GeoJournal* **72:**137–148

Foody GM, See L, Fritz S, van der Velde M, Perger C, Schill C and Boyd DS. 2013. Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project, *Transactions in GIS*, **17:** 847-860. doi: 10.1111/tgis.12033

Foody GM, See L, Fritz S, van der Velde M, Perger C, Schill C, Boyd DS and Comber A. 2014. Accurate attribute mapping from volunteered geographic information: issues of volunteer quantity and quality, *The Cartographic Journal*, (in press).

Freberg K, Saling K, Vidoloff KG, Eosco G. 2013. Using value modeling to evaluate social media messages: The case of Hurricane Irene, *Public Relations Review*, **39**: 185–192

GCOS. 2010: Implementation plan for the global observing system for climate in support of the UNFCC (2010 update), GOOS-184, GTOS-76, WMO-TD/No. 1523) [www] http://www.wmo.int/pages/prog/gcos/Publications/gcos-138.pdf

Girres JF and Touya G. 2010. Quality Assessment of the French OpenStreetMap Dataset, *Transactions in GIS*, 2010, **14**(4): 435–459

Gonzales JJ. 2011. The Impact of the Internet of Things on Business and Society [www] http://www.fundacionbankinter.org/system/documents/8194/original/Chapter_4_The_impact_of_the_IoT_in_Business_and_society.pdf (accessed March 2014)

Goodchild MF. 2007. Citizens as sensors: The world of volunteered geography, *GeoJournal*, **69**(4):211-221.

Goodchild MF and Li L. 2012. Assuring the quality of volunteered geographic information, *Spatial Statistics*, **1**: 110-120.

Goodchild MF and Glennon JA. 2010 Crowdsourcing geographic information for disaster response: a research frontier, *International Journal of Digital Earth*, **3**(3): 231-241,

Goodwin S. 2013. Raspberry Pi, In: *Smart Home Automation with Linux and Raspberry Pi*, Apress publishers, pp 275-296, doi: 10.1007/978-1-4302-5888-9_8

Gura T. 2013. Citizen Science: Amateur experts, *Nature*, **496:** 259-261

Haberlandt U and Sester M. 2010. Areal rainfall estimation using moving cars as rain gauges **-** a modelling study, *Hydrol. Earth Syst. Sci*. **14**: 1139-1151

Haklay M, Basiouka S, Antoniou V and 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**: 315–322.

Heggestuen J. 2013. One In Every 5 People In The World Own A Smartphone, One In Every

17 Own A Tablet [www] http://www.businessinsider.com/smartphone-and-tablet-penetration-2013-10 (Accessed January 2014)

Heipke C. 2010. Crowdsourcing geospatial data, *ISPRS Journal of Photogrammetry and Remote Sensing* **65**: 550–557

Ho IW-H, Leung KK, Polak JW. 2009. Connectivity Dynamics for Vehicular Ad-hoc Networks in Signalized Road Systems, *21st International Teletraffic Congress 2009*, 15-17 Sept. 2009, Paris, France

Hochachka WM, Fink D, Hutchinson RA, Sheldon D, Wong WK and Kelling S. 2012 Data-intensive science applied to broad-scale citizen science. *Trends in Ecology and Evolution* **27**:130-137.

Holland G. 2008.“Will the real smart city please stand up?”, *Cities*, **12(**3): 303 - 320

Honicky RJ, Brewer EA, Paulos E, White RM. 2008. N-SMARTS: Networked Suite of Mobile Atmospheric Real-time Sensors, *NSDR’08,* August 18, 2008, Seattle, Washington, USA

Howe, J. 2006. Crowdsourcing: a definition. Wired Blog Network: Crowdsourcing [www] http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html (Accessed February 2014)

Hut R, Jong S, Giesen N. 2014. Using umbrellas as mobile rain gauges: prototype demonstration, *Geophysical Research Abstracts*, EGU2014-16418

Illingworth SM, Muller CL, Graves R, Chapman L. 2014. UK Citizen Rainfall Network: A pilot study, *Weather*, in press

IPCC, 2012. Summary for Policymakers. In: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 1-19

Johnston S. 2013. WeatherSignal: Big Data Meets Forecasting [www] http://blogs.scientificamerican.com/guest-blog/2013/10/11/weathersignal-big-data-meets-forecasting/ (Accessed March 2014)

Kazai G, Kamps J, and Milic-Frayling N. 2013. An analysis of human factors and label accuracy in crowdsourcing relevance judgements, *Information Retrieval*, **16**: 1-41.

Kidd C, Huffman G, Kirschbaum D, Skofronick-Jackson P, Joe P, Muller CL. 2014. So, how much of the Earth's surface is covered by rain gauges? *Geophysical Research Abstracts*, EGU2014-10300

Krause A, Mellon C, Horvitz E, Kansal A, Zhao F. 2008. Toward Community Sensing, *International Conference on Information Processing in Sensor Networks (IPSN),* 22-24 April 2008, St. Louis, Missouri, USA

Leijnse H, Uijlenhoet R., and Stricker JNM 2007. Rainfall measurement using radio links from cellular communication networks, *Water Resour. Res.*, **43**: W03201, doi:10.1029/2006WR005631.

Lorenz C, Kunstmann H. 2012. The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis. *J Hydrometeorol* **13**(5):1397–1420.

Madaus LE, Hakim GJ, Mass CF. 2014. Utility of dense pressure observations for improving mesoscale analyses and forecasts., *Mon. Wea. Rev*. doi:10.1175/MWR-D-13-00269.1

Mahoney B, Drobot S, Pisano P, McKeever B, O'Sullivan J. 2010. Vehicles as Mobile Weather Observation Systems, *Bulletin of the American Meteorological Society*, **91**: 1179–1182.

Mahoney WP. and O'Sullivan JM. 2012. Realizing the potential of vehicle-based observations, *Bulletin of the American Meteorological Society*, **94:** 1007–1018

Mass C. 2013. Critical Aircraft Weather Data Unused By NOAA [www] http://cliffmass.blogspot.co.uk/2013/09/critical-aircraft-weather-data-unused.html (Accessed March 2014)

Melhuish E, Pedder M. 2012. Observing an urban heat island by bicycle, *Weather*, **53**(4) 121-128

Messer H, Zinevich A, and Alpert P. 2006. Environmental monitoring by wireless communication networks, *Science*, 3**12**: 713.

Minda H and Tsuda N. 2012. Low-Cost Laser Disdrometer with the Capability of Hydrometeor Imaging, *IEEJ Tran*, **7**(S1): S132–S138, DOI:10.1002/tee.21827

Montjoye Y-A, Hidalgo CA, Verleysen M, Blondel VD. 2012. Unique in the Crowd: The privacy bounds of human mobility, *Scientific Reports* **3**: 1376, doi:10.1038/srep01376

Moore G. 1965. Cramming More Components onto Integrated Circuits, *Electronics Magazine*, **38** (8)

Muller CL. 2013. Mapping snow depth across the West Midlands using social media-generated data. *Weather*, **68**(3): 82.

Muller CL, Chapman L, Ferranti EJS, and Young DT. 2014. Crowdsourcing temperatures from vehicle sensors: Lessons from a high-resolution citizen science experiment, *Weather*, under review

Muller CL, Chapman L, Grimmond, CSB, Young DT, Cai X. 2013a. Sensors and the city: a review of urban meteorological networks. *International Journal of Climatology*, **33**(7): 1585–1600.

Muller CL, Chapman L, Grimmond, CSB, Young DT, Cai X. 2013b. Toward a Standardized Metadata Protocol for Urban Meteorological Networks. *Bulletin of the American Meteorological Society*, **94**(8): 1161–1185.

Neirotti P, De Marco P, Cagliano AC, Mangano G, Scorrano F. 2014. Current trends in Smart City initiatives: Some stylised facts, *Cities*, **38:** 25–36

Neis P, Zielstra D and Zipf A. 2012. The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011, *Future Internet*, **4**: 1-21

Nickinson P. 2013 Samsung Galaxy S4 specs [www] http://www.androidcentral.com/samsung-galaxy-s4-specs (accessed Dec 2013)

Nielsen M. 2011. *Reinventing Discovery: The New Era of Networked Science*, Princeton University Press

Nissenbaum H. 2004. Privacy as a contextual integrity, *Washington Law Review*, **79** (1), 101-139.

Normandeau K. 2013. Beyond Volume, Variety and Velocity is the Issue of Big Data Veracity [www] http://inside-bigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/ (Accessed November 2013)

NRC. 2007. *Successful response starts with a map: improving geospatial support form disaster management*. Washington, DC: National Academies Press

Ochoa-Rodriguez S, Rico-Ramirez MA, Jewell SA, Schellart ANA, Wang L, Onof C and Maksimovic C. 2013. Improving rainfall nowcasting and urban runoff forecasting through dynamic radar-raingauge rainfall adjustment, *7th International Conference on Sewer Processes and Networks*, 28-30 August 2013, Sheffield, UK

Oke TR. 2004. Siting and Exposure of meteorological instruments at urban sites, 27th NATO/CCMS International Technical Meeting on Air Pollution Modelling and its Application, Banff, October 25–29, 2004.

Overeem A, Leijnse H, Uijlenhoet R. 2013b. Country-wide rainfall maps from cellular communication networks, Proceedings of the National Academy of Sciences of the United States of America, 110: 2741-2745

Overeem A, Robinson JCR, Leijnse H, Steeneveld GJ, Horn BKP, Uijlenhoet R. 2013a. Crowdsourcing urban air temperatures from smartphone battery temperatures. *Geophysical Research Letters*, **40** (15), 4081-4085

Pal M and Foody GM. 2010. Feature selection for classification of hyperspectral data by SVM, IEEE Transactions on Geoscience and Remote Sensing, 48: 2297-2307. Pham H. 2014. Data Driven Education, *Adjacent Local Government* [www] http://www.adjacentgovernment.co.uk/lg-edition-001/data-driven-education/ (Accessed February 2014)

Rabiei E, Haberlandt U, Sester M and Fitzner D. 2013. Rainfall estimation using moving cars as rain gauges – laboratory experiments, *Hydrol. Earth Syst. Sci.*, **17**: 4701–4712

Rada EC, Ragazzi M, Brini M, Marmo L, Zambelli P, Chelodi M. 2012. Perspectives of low-cost sensors adoption for air quality monitoring, *U.P.B. Sci. Bull., Series D*, **74**(2): 1454-2358

Raykar VC and Yu S. 2011. An entropic score to rank annotators for crowdsourced labelling tasks, *Proceedings Third National Conference on Computer Vision*, Pattern Recognition, Image Processing and Graphics, IEEE, 29-32

Raykar VC and Yu S. 2012. Eliminating spammers and ranking annotators for crowdsourced labelling tasks, *Journal of Machine Learning Research*, **13**: 491-518.

Raymond ES. 2001. *The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, Revised Edition, Publ. O'Reilly Media Inc.

Robinson J. 2013. What Google Activity Recognition means for apps [www] http://developer.android.com/training/location/activity-recognition.html (Accessed March 2014)

Roy HE, Pocock MJO, Preston CD, Roy DB, Savage J, Tweddle JC, Robinson LD. 2012. *Understanding Citizen Science and Environmental Monitoring*, Final report on behalf of UK-EOF. NERC Centre for Ecology and Hydrology and Natural History Museum [www] http://www.ceh.ac.uk/products/publications/documents/citizensciencereview.pdf (Accessed March 2013)

Scanfeld D, Scanfeld V, Larson EL 2010. Dissemination of health information through social networks: twitter and antibiotics, *Am J Infect Control*, **38(**3):182-8

Schaller RR. 1997. Moore's Law: past, present and future, *IEEE Spectrum*, **34** (6): 52-59

See L, Comber A, Salk C, Fritz S, van der Velde M, Perger C, Schill C, McCallum I, Kraxner F, Obersteiner M. 2013. Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts, *PLoS ONE* **8**(7): e69958. doi:10.1371/journal.pone.0069958

Shan Q and Brown D. 2005. Wireless Temperature Sensor Using Bluetooth, *IWWAN 2005: International Workshop on Wireless Ad Hoc Networks*, London, 23-26 May 2005

Sheehy J. 2013. Crowdsourced Weather Forecasting with Google Glass [www] http://www.jacobsheehy.com/2013/05/weather-forecasting-with-google-glass (accessed Dec 2013)

Silvertown J. 2009. A new dawn for citizen science. *Trends Ecol. Evol*. **1118**:1–5

Snow R, O'Connor B, Jurafsky D and Ng AY. 2008. Cheap and fast – but is it good? Evaluating non-expert annotations for natural language tasks, *Proceedings 2008 Conference on Empirical Methods in Natural Language Processing*, 25-27 October, Hawaii, 254-263.

Steeneveld, G. J., S. Koopmans, B. G. Heusinkveld, L. W. A. van Hove, and A. A. M. Holtslag. 2011. Quantifying urban heat island effects and human comfort for cities of variable size and urban morphology in the Netherlands, *J. Geophys. Res.*, 116, D20129, doi:10.1029/2011JD015988.

Sui DZ, Elwood S and Goodchild MF. 2013. *Crowdsourcing Geographic Knowledge*. Berlin: Springer.

Tahmo. 2013. Trans-African Hydro-Meteorological Observatory [www] www.tahmo.org/ (Accessed January 2013)

Thornton T. 2013. TeamSurv - using the crowd to monitor the oceans, *UK-IMON International Workshop on New Monitoring Technologie*s, 10-12 September 2013, Southampton UK

Tomlinson CJ, Chapman L, Thornes JE, Baker CJ. 2011. Including the urban heat island in spatial heat health risk assessment strategies: a case study for Birmingham, UK, *International Journal of Health Geographics*, **10:**42

Tomlinson CJ, Prieto-Lopez T, Bassett R, Chapman L, Cai X-M, Thornes JE, Baker CJ.

2013. Showcasing urban heat island work in Birmingham – measuring, monitoring, modelling and more, *Weather*, **68**(2), 44049

Trumbull D, Bonney R, Bascom D, and Cabral A. 2000. Thinking scientifically during participation in a citizen-science project, *Science Education,* **84** (2): 265–275

Tweddle JC, Robinson LD, Pocock MJO, Roy HE. 2012. Guide to citizen science: developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK, Natural History Museum and NERC Centre for Ecology and Hydrology for UK-EOF [www] www.ukeof.org.uk (Accessed January 2014)

UN. 2009. World urbanization prospects, The 2007 Revision Population Database. [www] http://esa.un.org/unup/.

Walsh D. 2012. The tricky business of counting rain. NY Times [www] http://green.blogs.nytimes.com/2012/07/02/do-not-publish-the-tricky-business-of-counting-rain (Accessed January 2013).

Wechsler D. 2014. Crowdsourcing as a method of transdisciplinary research – tapping the full potential of participants, *Futures*, http://dx.doi.org/10.1016/j.futures.2014.02.005

Welinder P, Branson S, Belongie S and Perona P. 2010. 'The multi-dimensional wisdom of crowds', Advances in Neural Information Processing Systems 23, (edited by John Lafferty, Chris Williams, John Shawe-Taylor, Richard Zemel and Aron Culotta), Curran Associates, NY, 2424-2432.

Wiersma YF. 2010. Birding 2.0: Citizen Science and Effective Monitoring in the Web 2.0 World, *Avian Conservation and Ecology,* **5**(2): 13

Wiggins A, Crowston K. 2011. From conservation to crowdsourcing: A typology of citizen science. In System Sciences (HICSS), 2011 44th Hawaii International Conference on System Science,1-10. IEEE.

Willems P, Arnbjerg-Nielsen K, Olsson J, Nguyen VTV. 2012. Climate change impact assessment on urban rainfall extremes and urban drainage: Methods and shortcomings, *Atmospheric Research,* **103**: 106–118

Wolters D and Brandsma T. 2012. Estimating the Urban Heat Island in Residential Areas in the Netherlands Using Observations by Weather Amateurs. *J. Appl. Meteor. Climatol.,* **51**: 711–721.

Woodruff SD, Slutz RJ, Jenne RL, Steurer PM. 1987. A comprehensive ocean-atmosphere data set, *Bull. Am. Meteorol. Soc*. **68** (10): 1239-1252

Worley SJ, Woodruff SD, Reynolds RW, Lubker SJ, Lott N. 2005. ICOADS release 2.1 Data and Products, *International Journal of Climatology,* **25**:823-842

Yatagai A, Kamiguchi K, Arakawa O, Hamada A, Yasutomi N, Kitoh A. 2012.

APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull Am Meteorol Soc* 9**3**(9):1401–1415.

Young DT, Chapman L, Muller CL, Cai XM, Grimmond, CSB. 2014. A low-cost wireless temperature sensor: evaluation for use in environmental applications, *J. Atmos. Ocean. Tech., in press.*

Tables

Table 1: Examples of current atmosphere, weather and climate-related crowdsourcing projects and techniques

Project	Type	Data	Summary	Reference/URL
UKSnowMap	Web 2.0, citizen science	Snow rating, location	UK citizens tweet a snow rating (out of 10) which are shown on map	http://uksnowmap.com/
Snow Tweets	Web 2.0, citizen science	Snow depths, location	World-wide citizens tweet snow depths which are shown on map?	http://www.snowtweets.org
CoCoRaHS	Web 2.0, citizen science, amateur weather stations	Rainfall amount, location	US citizens upload information about precipitation amount as measured by manual gauges	http://www.cocorahs.org/Cifelli *et al.*, 2005
UCRaiN	Web 2.0, citizen science, amateur weather stations	Rainfall amount, location	UK citizens upload information about precipitation amount as measured by manual, home-made gauges	Illingworth *et al.* (2014)
Global Learning and Observations to Benefit the Environment (GLOBE)	Citizen science, amateur weather stations and other environmental sensors	A range of environmental data , inc. weather data	The GLOBE Programme is an established, international science and education project whereby students and teachers can take scientifically valid environmental measurements and report them to a publicly available database.	www.globe.gov/ Finarelli (1998)
WeatherSignal	Smart device, mobile app	Location, temperature, pressure, humidity, weather reports, acceleration, magnetic flux, light	A mobile phone application for obtaining weather data from mobile phone users	http://weathersignal.com/
PressureNet	Smart device, Mobile app	Pressure	App automatically collects atmospheric pressure measurements using barometers in Android devices.	http://pressurenet.cumulonimbus.ca/
Birmingham snow depth	Web 2.0, citizen science	Snow depth, location	Birmingham citizens tweet snow depths	Muller (2013)
City temperatures from smart phone battery temperatures	Smart device, mobile app	Mobile phone battery temperature; Air temperature proxy, location	Temperature data derived from smart phone batteries sensors (not specifically designed for crowdsouricng the weather) are fed into a heat transfer model to produce daily air temperatures averaged over a city.	Overeem *et al.* (2013); http://www.opensignal.com
IntelliDrive/Vehicle Data Translator	Vehicle sensors	Temperature, position	Data from vehicle sensors are obtained and processed	Drobot *et al.* (2009; 2010), Anderson *et al.* (2012)
Birmingham car	Web 2.0, citizen	Air temperature, location	Birmingham citizens tweet car thermometer	Muller *et al.* (pers comms.)

temperatures	science, vehicle sensors		temperature readings	
Old Weather	Citizen science	Archive weather data	Citizens transcribe mid-19th century ship logs	http://www.oldweather.org/
OPAL contrail	Citizen science	Contrail length survey	UK citizens noted the length of any contrails they could see over a fixed campaign period for comparison with data at aircraft altitude.	http://www.opalexplorenature.org/climatesurvey
Cyclone Centre	Citizen science	Archive	Citizen scientists manually classifying 30 years of tropical cyclone satellite imagery.	http://www.cyclonecenter.org
TeamSurv	Ship sensors, Citizen science	Water depth and position	Mariners help create better charts of coastal waters by logging depth and position whilst at sea and uploading data to the web for processing and display.	http://www.teamsurv.eu/
Precipitation Intensity Near the Ground (PING) / meteorological Phenomenon Identification Near the Ground (mPING)	Citizen science	Rainfall amount, rainfall type, location	Citizens upload information about precipitation amount and type, as well as the type of weather that is occurring	Binau, 2012 Elmore *et al.*, 2014 http://www.nssl.noaa.gov/projects/ping/
European Severe Weather Database	Citizen Science	Tornados, severe wind, large hail, heavy rain, funnel cloud, gustnado, dust devil, heavy snowfall / snowstorm, ice accumulation, avalanche, damaging lightning	Eye-witness reports and mapping of severe weather across Europe	http://www.essl.org/cgi-bin/eswd/eswd.cgi
UK Storm 2013 crowdmap	Web 2.0, citizen science	Location, information about storm damage	Map showing location and storm-related updates	https://ukstorm2013.crowdmap.com/
Twitcident	Web 2.0, citizen science	Geo-located information about a range of hazards / emergency incidents	Tweeted information for a range of applications in the public safety domain.	http://www.twitcident.org
Air Quality Egg	Citizen science, amateur weather stations	NO2, CO, temperature, humidity	Low-cost, WiFi-enabled air quality sensor	http://airqualityegg.com/
IBM Deep Thunder	Amateur weather stations	Range of weather data	Targeted weather forecasting program providing minute-by-minute, highly localized forecasts, using a combination of public weather data from NOAA, NASA, the U.S. Geological Survey, WeatherBug, and	http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepthunder/

			other weather sensors.	
Metwit	Mobile app, citizen science	Weather conditions	Real-time weather information via smart app	https://metwit.com/
UK Met Office ‘Weather Observation Website’ (WOW)	Amateur weather stations	Range of weather data and metadata	Amateur weather observers website for visualising data (including metadata and quality flags)	Bell *et al.* (2012) Tweddle *et al.* (2012) http://wow.metoffice.gov.uk
Meteoclimatic	Amateur weather stations	Range of weather data and metadata	A large real-time network of amateur automatic weather stations covering the Iberian Peninsula	http://www.meteoclimatic.com/
Weather Underground	Amateur weather stations	Range of weather data	Amateur weather observers website for archived data	http://www.wunderground.com/personal-weather-station/signup
Citizen Weather Observer program (CWOP)	Amateur weather stations	Range of weather data	Amateur weather observers website for archived data	http://www.wxqa.com
Weather Bike	Bicycle platform, Amateur weather stations	Location, temperature, wind	Low-cost sensors attached to a bicycle	Cassano (2013)
AirPi	Low-cost sensors	Temperature, humidity, air pressure, light levels, UV levels, carbon monoxide, nitrogen dioxide, smoke level	A Raspberry Pi shield kit that can record a range of data and upload to the internet	http://airpi.es/
Measuring rain using microwave links from cellular communication networks	Hidden networks	Rain	Utilising received signal level data from microwave links in cellular communication networks to monitor rainfall	e.g. Messer *et al.* (2006), Leijnse *et al.*, (2007), Overeem *et al.* (2013b)

1 Table 2: Potential uses and applications of a variety of crowdsourced data

Application	Examples of crowdsourced data type	Examples of potential uses
High-resolution, localised observations	• Sensor data from mobiles, vehicles, trains, bikes (e.g. GPS, signal, other sensor and proxy data) • Smart meters in homes and offices • Citizen science and web 2.0	Tracking thunder and lightning, tornadoes, hurricanes; monitoring, forecasting and managing flooding; heatwaves; air pollution events; societal applications (e.g. health, infrastructure management, city-planning, risk assessment)
Decision-making	• All types	Real-time, high spatiotemporal to inform decision-making for planning, adaptation, mitigation, management
Risk Assessment	• Low-cost citizen sensors and weather stations • Smart phone sensors • Citizen science data	Better monitoring and assessment of hazard risks and vulnerabilities.
Modelling	• Low-cost citizen sensors and weather stations • Smart phone sensors • Citizen science data	Higher resolution data for model evaluation and assimilation
Forecasting/nowcasting	• Low-cost citizen sensors and weather stations, mobile phone sensors, citizen science data	Higher resolution data than standard in situ measurements; use of real-time data
Ground-truth remote sensing data (satellite, radar)	• Low-cost/citizen measurements of rainfall, air quality, snow etc	Increase data-availability in data sparse areas (e.g. low-income countries, less-accessible areas); Improve retrieval algorithms; Production of new combined data products.
Scientific research	• All types	Higher spatiotemporal data could provide new scientific insights where data is lacking
Climate monitoring	• All types	Higher spatiotemporal observations for long-term climate monitoring, particularly useful for exploring variability in morphologically heterogeneous areas such as cities
Infrastructure (e.g. roads, rails, cycle paths, pedestrian routes, energy, ICT)	• Sensor data from mobiles, vehicles, trains, bikes (e.g. GPS, signal, other sensor and proxy data) • Smart meters in homes and offices • Mobile/WiFi signal strength	Real-time, high spatiotemporal to inform decision-making, re-routing traffic, informing gritting routes, clearing gutters during flash flooding, better control of energy use, understanding resilience of networks under different weather conditions.
Emergency services (fire; police; hospitals/ambulance)	• Sensor data from mobiles, vehicles, trains, bikes (e.g. GPS, signal, other sensor and proxy data) • Smart meters in homes and offices • Citizen science and web 2.0	Could assist with predicting/identifying areas at risk (e.g. anti-social behaviour, thefts, illness during heatwaves, road accidents, illness caused by snow/ice/flood)
Health	• Sensor data from mobiles, vehicles, trains, bikes (e.g. GPS, signal, other sensor and proxy data) • Smart meters in homes and offices	Predicting/identifying patterns during outbreaks and identifying areas at risk (e.g. seasonal illness such as hay fever, disease outbreaks, accidents and illness during extreme events)

	• Citizen science and web 2.0	
Agriculture	• Low-cost citizen sensors and weather stations	Monitoring of annual and seasonal variability for economic and production applications; microscale variability across small geographic areas (e.g. soil moisture) for increasing productivity.
Insurance and post-event analysis	• Low-cost/citizen measurements of rainfall, air quality, snow etc • Citizen science and web 2.0	For example, identifying flood damage; flood depth/occurrence; advising appropriate engineering solutions
Knowledge transfer – private / public sector use	• All types	More open, cost-effective data for use in industrial applications
Public engagement / science communication	• All types, particularly citizen science and web 2.0	Engages people with their local neighbourhood and involves them in science/data applications for public benefit
Education	• All types, particularly citizen science and low-cost sensors	More data for use in education, without the need for expensive equipment; engaging students with scientific research; encouraging science, technology, engineering, mathematics (STEM) uptake

List of Figures

Figure 1: Venn diagram showing the interaction of animate and inanimate crowdsourcing components, including active and passive techniques.

Figure 2: Estimation of air temperature from smartphone battery temperatures: comparison with data from (top) WMO Birmingham airport site (located just outside the city) and (bottom) two central Birmingham UKMO sites (which are located in the vicinity of a large number of battery readings): (a) Map of Birmingham (UK; ©OpenStreetMap contributors; openstreetmap.org) showing locations of selected smartphone battery temperature readings (blue dots) from 1st June to 31st August 2013 and location of WMO and UKMO weather stations (red ovals) (b) Time series of daily averaged observed and estimated air temperatures, as well as battery temperatures in Birmingham for same period. (c) Scatter plot of estimated daily air temperatures against observed daily air temperatures based on data from Birmingham for1st June to 31st August 2013. Grey line is y = x line. ME denotes mean error (bias), MAE is mean absolute error, CV is coefficient of variation, ρ^2 is coefficient of determination. CAL and VAL stand for calibration and validation data set, respectively. WMO nr. is World Meteorological Organization station index number.

Figure 3: Map showing the sparse global distribution of stations included in the Monthly Climate Data for the World report for July 2013 (Source: NOAA National Climatic Data Centre, http://www1.ncdc.noaa.gov/pub/data/mcdw/)

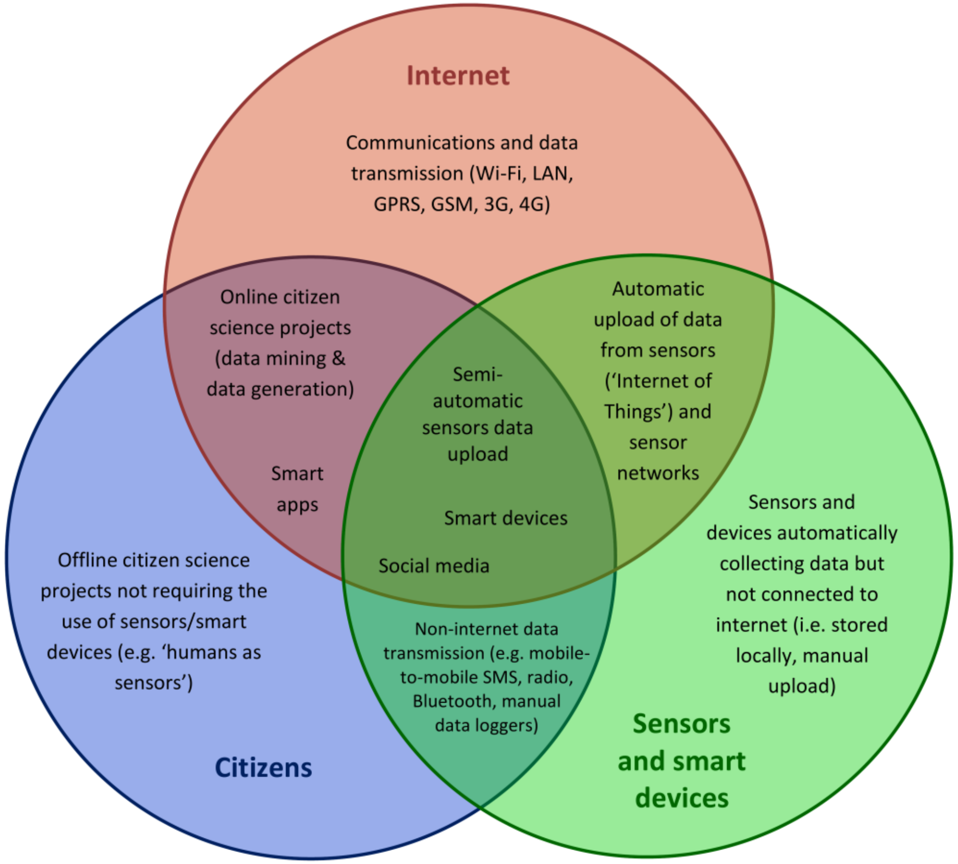
Internet
Communications and data transmission (Wi-Fi, LAN, GPRS, GSM, 3G, 4G)
Online citizen science projects (data mining & data generation)
Automatic upload of data from sensors (‘Internet of Things’) and sensor networks
Semi-automatic sensors data upload
Smart apps
Smart devices
Social media
Offline citizen science projects not requiring the use of sensors/smart devices (e.g. ‘humans as sensors’)
Sensors and devices automatically collecting data but not connected to internet (i.e. stored locally, manual upload)
Non-internet data transmission (e.g. mobile-to-mobile SMS, radio, Bluetooth, manual data loggers)
Citizens
Sensors and smart devices

Figure 1

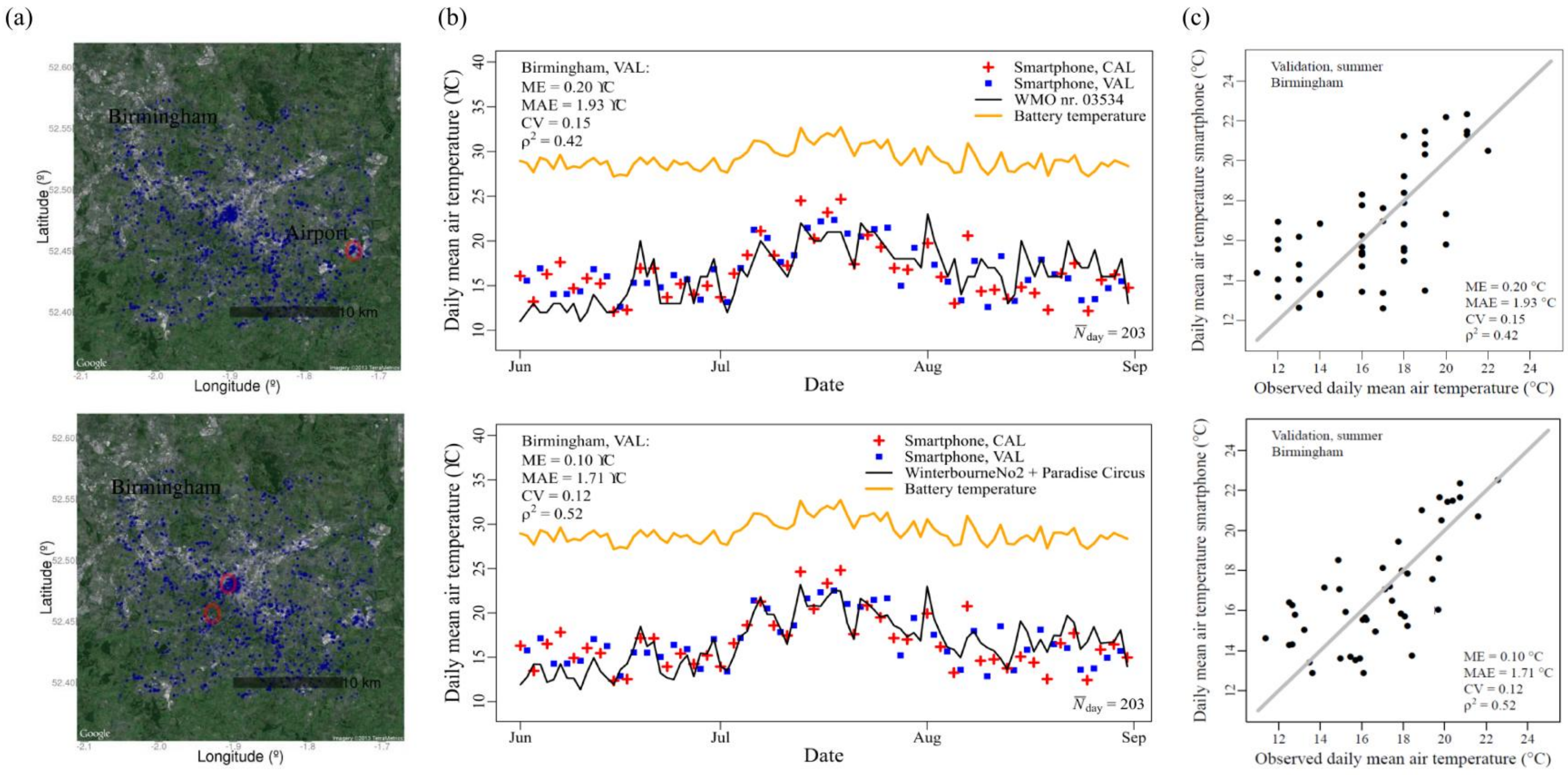
(a)
(b)
(c)
Birmingham
Airport
10 km
Latitude (º)
Longitude (º)
Google
Birmingham, VAL:
ME = 0.20 °C
MAE = 1.93 °C
CV = 0.15
ρ² = 0.42
Smartphone, CAL
Smartphone, VAL
WMO nr. 03534
Battery temperature
$\overline{N}_{day}$ = 203
Daily mean air temperature (°C)
Jun
Jul
Aug
Sep
Date
Validation, summer
Birmingham
ME = 0.20 °C
MAE = 1.93 °C
CV = 0.15
ρ² = 0.42
Daily mean air temperature smartphone (°C)
Observed daily mean air temperature (°C)
Birmingham
10 km
Latitude (º)
Longitude (º)
Google
Birmingham, VAL:
ME = 0.10 °C
MAE = 1.71 °C
CV = 0.12
ρ² = 0.52
Smartphone, CAL
Smartphone, VAL
WinterbourneNo2 + Paradise Circus
Battery temperature
$\overline{N}_{day}$ = 203
Daily mean air temperature (°C)
Jun
Jul
Aug
Sep
Date
Validation, summer
Birmingham
ME = 0.10 °C
MAE = 1.71 °C
CV = 0.12
ρ² = 0.52
Daily mean air temperature smartphone (°C)
Observed daily mean air temperature (°C)

Figure 2

Figure 3