tecnalia Inspiring Business

www.tecnalia.com

NEW "BIG-DATA" SOURCES FOR METEO-CLIMATIC COMMUNITY

Santiago Gaztelumendi [1,2] and José Daniel Goméz de Segura [1].

1- Meteorology Area, Energy and Environment Division, TECNALIA R&I.

2- Basque Meteorology Agency (Euskalmet).

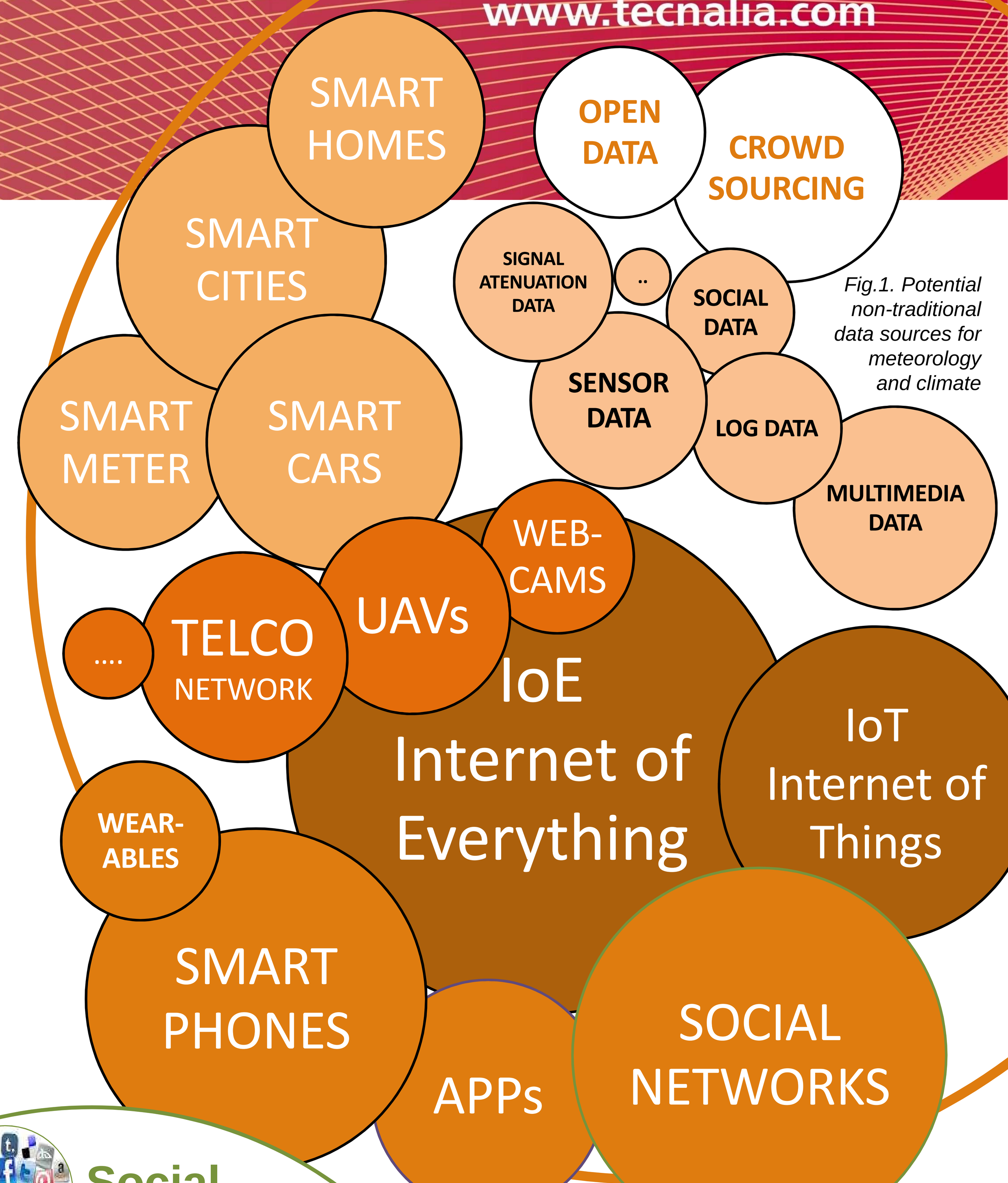


Fig.1. Potential non-traditional data sources for meteorology and climate

Abstract

Traditional operational meteorology relies on big amount of data, from different simulation models, observation sources and different human-based products. This classical scenario involves large amounts of more or less structured data. Nowadays, connected sensors are becoming ubiquitous and social networks offer valuable real-time geolocalised information. Big data may offer a greater insight and result in better and new products for end-users. In this context, we understand "big data" as data that is too complex to be processed by traditional means in an acceptable way. The complexity comes from the multiple dimensions of data and particularly from Volume, Variety, Veracity or Velocity, the 'so called four "Vs" of big data.

In this work we focus on "Variety". We present some ideas around non-traditional data sources available today or in near future for meteorology and climate community. We analyze different promising data sources as a first step in order to try to incorporate them in the value chain of the climate and meteo business. We present some examples and we try to imagine in which ways they can be useful for meteo-climatic community. Finally, some conclusions and remarks about potential problems associated for its operational use are presented.

Introduction

Thanks to Big Data (BD) and Information and Communication Technologies (ICT), there are multiple possibilities for incorporating new data acquisition system in to the meteo-climatic chain from non-traditional sources (Gaztelumendi et al 2015). The extension of the open-data philosophy by many public administrations in different parts of the World, and particularly the application of cross-border directives, such as INSPIRE in Europe, foster a growing availability of data. Moreover, the emergence of different data-sources that were not available in the 20th Century, with bigger momentum based on technological drivers such as mobile phone network, geolocation, electronic component cost reduction or wide-spreading of social networks, make necessary an analysis about the usability of these new information sources and their potential utility in the operational meteorology and climatology fields.

Many forms of non-traditional (data) has the potential to be used in 21st century meteorology and climate business (see fig 1). Here we focus on a particularly promising area around internet connected smartphones and crowdsourcing, where mobile phone sensors, wearables and social networks seems to be a huge contribution to big data "variety" flavor with vast applications including the weather business.

Methodology

This work is based on Technology Watching procedures for detection, analysis and reporting innovative behavior trends (Tecnalia 2014). These methodologies are applied in order to take full awareness of technological competences that will be a reality in a more or less near future. In the process of search and detection different data sources are used including patents, research groups, scientific publications repositories, technology offers and demands, international conferences, symposiums, market trends, grant calls, innovative projects, market trends, blogs, google, etc.

Here we present some results extracted from the application of such methodologies to the field of meteorology and climate and particularly to non-traditional data sources.

Discussion

Mobile phone Sensors

Today smartphones has plenty of sensors for different purposes . Although at present these are more commonly found in higher-end phones, the progressive addition of sensors in all kind mobile devices is a increasing tendency for the future (e.g. Fenghua Li et al 2015).

The most obvious are all about movement. Like the accelerometer, for measuring movement and orientation, or the gyroscope, for measuring angular rotation.. But there are also many environmental sensors, which measure things like temperature, barometric pressure, light, etc. By instance last Samsung generation includes : Fingerprint scanner, Heartrate, RGB Ambient Light, Relative humidity, Env .Temperature, Barometer, NFC, Gyroscope, Accelerometer, Bluetooth radio, WiFi radio, FM radio, Cell radio, Front camera, Rear camera, GPS, Magnetic field, Light flux, Battery temp, Touch and Microphone (Samsung 2016).

Here the question is how a mobile phone (or connected wearable) is able to act as a kind of AWS (Automatic Weather Station) collecting directly or indirectly different type of meteo-climatic information by different ways.

Wearables

Smartwatches with connectivity that includes altimeter, barometer, thermometer, gyroscope, accelerometer and humidity sensors can be used as an alternative data source (Apple watch, Microsoft band, Samsung gear, etc..)
Fitness sport oriented.
healthcare & medical. wearables
Smart cloths, eyewear, skin patches also has some meteo-climatic potential.

Connectivity and Crowdsourcing

Meteo-climatic oriented readings of a single smartphone's usually can not be considered directly due to lack of representativeness and quality. But taking readings from many devices, it is possible to extract reliable and accurate measurements of different ambient parameters. The fact that collected data, from mobile phone sensors network, are spatially and temporally correlated, is exploited to better assess their correctness.
Data collection can be passive with citizens' equipment automatically submitting data to a database without any specific user actions necessary to acquire the data. Even totally implicit with users do not necessarily knowing they are contributing. Or can be explicit and active, with different degree on participation in the measurement process where participants following a formal procedure. Participatory-sensing (implicit or explicit crowdsourcing) in a world with billions of mobile phones, poses plenty of challenges and opportunities for big data and meteo-climatic communities.

Social Networks

Social media offer a quick and easy way for the public to real time interaction from their own locations. Users can post, in different social media tools, severe weather descriptions, local detailed information, incidences, photos, videos etc., usually from mobile phones when and where internet connection are available.

Since the beginnings, social media has been considered as a potential source of valuable meteorological information for different purposes (e.g. Hyvarinen et al 2010). Today with nearly 1400 millions active users on Facebook and hundreds of millions on other networks such as Twitter, Google +, Instagram, flykr, youtube, etc the key question is how to extract the meteo-climatic value, that sure is there, from this huge amount of data from all around the world. Again and as usually in crowdsourcing world, different approaches are possible depending of the degree of compromise and effort that is required by the public.

Facebook and Twiter are already a widespread communication channel in different National Weather Services all around the world (Gaztelumendi et al 2015, Palacio et al 2016). NMHSs develop effective relationships with their audience through constructive, mutually beneficial online dialogue. Unless social media are recommended if an adequate ongoing resources to maintain the level of interactivity desired with users are possible (Gill et al 2012), at the end they are mostly used for one-way communication And particularly extended in severe weather real time communication as is the case of Euskalmet (Gaztelumendi et al 2012).

Specific sensors can be incorporated to the phone as an add-on or widget

Examples:
Vaavud is an add-ons that convert the smartphone in an anemometer. Model **Mjolnir** is based on two rotating cups along with internal magnets, attached to the headphone jack, the frequency of motion results in a periodic distortion of the magnetic field around the device, that is measured and Interpreted as wind speed. Model **Sleipnir** designed to capture wind speed and direction using two curved blades and the internal optic sensor (https://vaavud.com)
iSPEX, is a low-cost optical add-on for smartphones the smartphone camera transform the phone into a spectropolarimetric instrument, providing a direct measure of the sky polarization. In combination with corwdsourcing can provide aerosol optical thickness (AOT) high resolution maps at urban scale (Snik et al 2014).
Sundroid is a iniciative based in a small sensor unit (as a wearable) that measures the UV radiation transmiting via bluetooth solar radiation data to the smartphone and to an app (Fahrni et al 2011).
Sensodrome is smart portable sensor platforms that can turns a smartphone in a wearable, programmable, sensing computer. Packing more than 11 sensors into one tiny package, including a carbon monoxide detector, non-contact thermometer, gas leak detector, lux meter, weather station, diagnostic tool, etc.
Netatmo is a complete amateur home weather station that offer real-time insight from wind speed and expected rainfall to solar radiation and UV levels with fully connectivity and smartphone capabilities

Sensors originally there for different purposes can be used for meteo-climatic data acquisition after some creative steps

Examples:
Measurements from the **camera** (originally present to capture images) and supervised image classification can be used for characterization of cloud cover (e.g. Parisi et al 2016)
Measurements from **battery temperature sensors** (originally present control battery charging as well as prevent damage) can be used to measure outdoor ambient temperature (e.g. Overeem et al 2013, Pape et al 2015).
Measurements from **pressure sensor** (originally present to improve altitude data from GPS fixes) can be used to determine detailed pressure fields at mesoscale for high resolution weather prediction(e.g. Mass and Madaus 2014)
Measurements from **thermometer, hygrometer, barometer and light sensor** is used by weathersignal for atmospheric measurement of air temperature , humidity, pressure and sky conditions (overcast, sunny, misty conditions) (Weathersignal).
Measurements **from image sensor** with location, orientation and time data can be used for atmospheric turbidity estimations. By first calibrating the image radiometrically and then comparing the intensity with a physics-based model of sky luminance. In this original approximation the user takes a picture of the sky that is transferred to a back- end server for processing (Puduri et al 2010).
Measurements from the **camera** can be used for measurement of direct solar irradiance and aerosol optical depth at 340 nm and 380 nm can be obtained with an app collecting, calibrating, processing and calculating the data in the smartphone itself for locations that have low aerosol loading (Igoe 2014)

APPs

A very popular tool for explicit crowdsourcing of meteorological or environmental data are mobile apps.

Examples:
WeatherSignal (http://weathersignal.com/) is an app that allows users to monitor the many sensors on their device - light, magnetism, pressure and more - while contributing to a realtime weather map available to users. The final aim of the owners (Opensignal) is to have thousand people using the app and sending data in each city contributing to a high detail.
weather picture
Sunshine (https://thesunshine.co/). This popular app is based in weather experimented by the users and feelings transmitted by users to the community. Reports are processed in order to convert the information into weather predictions.
CrowdMag (http://www.ngdc.noaa.gov/geomag/crowdmag.shtm) is an app to share magnetic data. In this project , the geomagnetism group of NOAA's National Centers for Environmental Information (NCEI) explore whether digital magnetometers built in modern mobile smartphones can be used as scientific instruments. With CrowdMag mobile apps, phones all around the world send magnetometer data to a server and after processing make data available for its use
MyShake (http://myshake.berkeley.edu/) earthquake app uses cellphones to detect earthquakes as soon as they start. Accelerometer in the phone records earthquakes as they happen and broadcast the data to a central service
GPS-based geographical collaborative applications for **transportation and route information** as Waze (https://www.waze.com/es/) or Moovitap (http://moovitapp.com/), with over 30 million users worldwide contain incidences and advisory introduced by the users that can be useful.

The **implicit** approach,

is based on the fact that many social media messages tell about weather direct and indirectly, and that this information is particularly abundant during high-impact weather. Many internet services available today from smartphone has multimedia content (photos and video) that can also be used. Today many NMHSs and private companies are using implicit social media information for meteo-climatic purposes but just in a earlier stages of a promising world.

The **explicit** approach

is a sort of 21st century actualization of the concept of volunteers or "meteorological observer community". where we try to consolidate a online social weather community in order to obtain as many information as possible. Usually as a kind of weather report. This weather reports goes from non-real time manual detailed reports included in a web formulary, to specific real time information for a particular meteorological aspect that user introduce in an APP. Information can be numeric and continuously sent by an non-proffesional AWS , or be punctual including text, photo or video. Many examples of this new era of "e-weather observers" are available at different maturity stages, some examples

Twitter is an online social networking service that enables users (more than 320 millions) to send short 140-character messages with attached elements. Is a proven effective way to disseminate the goings on in the atmosphere rapidly and succinctly and has the ability to show video, annotate images, and quote other people. in our view today the best platform to exploit data for real time and nowcasting purposes. Twitter is used in NMHSs and private services directly or in combination with other tools for recollecting ground truth, evaluating text messages, monitoring hashtag, looking for keywords, analyzing traffic, extracting information from photos and videos, and others ways.

MeteoSwiss launched a pilot experiment collaborating with app users to collect a data set of hail observations for training and verification of radar measurements. (Noti et al 2016)

WOW project. The metoffice weather observation project for public data submitting using different equipment.. Manual and automatic observations can be introduced in the system, including "quick observations" with attached photo, or automatic observations from amateur stations. System is web based and observations introduced are not anonymous. (http://wow.metoffice.gov.uk/).

mPING project. The NOAA National Severe Storms Laboratory is collecting public weather reports through a free app available for smart phones or mobile devices. The app is called "mPING," for Meteorological Phenomena Identification Near the Ground. One of its main objectives is to obtain information about type of precipitation at ground in order to improve radar information. System is based on an app and allows to submit weather observations anonymously (http://mping.nssl.noaa.gov/).

SINOBAS project (http://sinobas.aemet.es/) This web based tool developed by Spanish AEMET, for reporting singular atmospheric observations, is a Volunteered Geographic Information (VGI) system, based on Google maps, where citizens help to building a data base of weather phenomena (Gutierrez et al 2015).

Weatherundergroud (https://www.wunderground.com) private company using observations from over 200,000 personal weather stations all around the world and supporting weather users communities

ESWD project (https://www.eswd.eu/). The European Severe Weather Database operated by the European Strom Laboratory is a collaborative project that includes networks of voluntary observers, meteorological services and general public providing detailed and quality-controlled information on severe convective storm events over Europe (Dotzek et al 2009). System is web based and not specific oriented to real-time.

....

.....

Youtube is today a video consumers community with over 1000 millions users, is used today n many NMHSs just for diffusion purposes and particularly for disseminating weather and climatic reports.

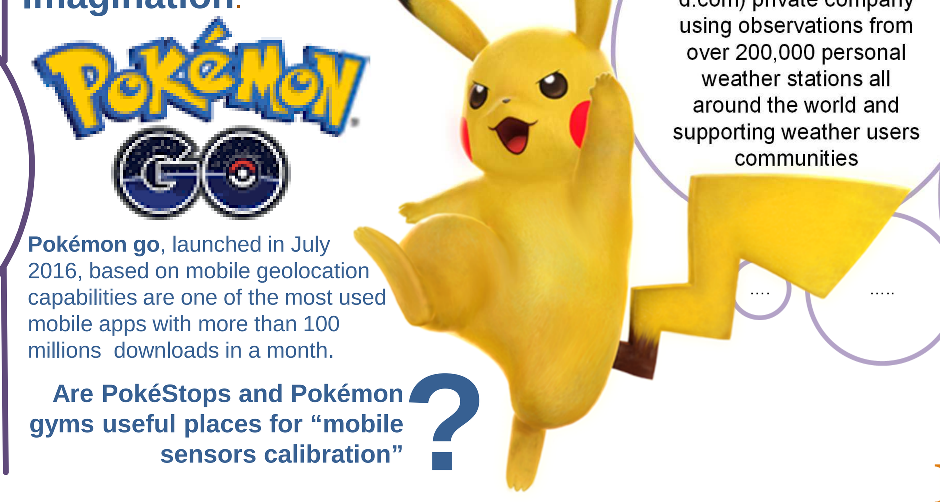
Filckr with more than 112 millions users is an online community where users share and embed personal photographs (1 million photos shared daily) with plenty of possibilities. (e.g.).

Instagram. It is an online mobile photo-sharing, video-sharing, and social networking service that enables its users to take pictures and videos, and share them either publicly or privately on the app, as well as through a variety of other social networking platforms. With more than 500 millions users sharing 80 millions of photos and videos each day and half of comments posted in the first 6 hours, possibilities are huge.

WhatsApp is a cross-platform instant messaging client for smartphones that uses the Internet to send text messages, documents, images, video, user location and audio messages to other users using standard cellular mobile numbers. Not a Social Network but has more than 1000 millions users sending 42.000 millions messages each day. They are just used in some NMHSs for internal communication.

Facebook the first ranking social network (more than 1100 millions users), allows exchange messages, post status updates and photos, share videos, etc. It is very popular but more difficult to build a good feed of real time weather information than Twitter. The recently launched app Facebook lite can be a good new opportunity. Today is mainly used in NMHSs for general communication purposes

A future with plenty of opportunities and challenges open for High Tec, Big Data and **Imagination**.



Pokémon go, launched in July 2016, based on mobile geolocation capabilities are one of the most used mobile apps with more than 100 millions downloads in a month.

Are PokéStops and Pokémon gyms useful places for "mobile sensors calibration"?

Conclussions & Remarks

The exponential growth of smartphone use, their ubiquity, the social use, and the sensors that are integrated in them can be of **great value if adequately processed**.

Usually single measurements and recollected data are not quality enough, but the **vast number** makes feasible to define algorithms for data pruning based on redundancy of observed values.

Mobile crowdsourcing and the exploitation of social media data can lead to **veracity** problems, an **urban bias**, as well as **safety** and **privacy** concerns.

Applications in meteorology involves activities totally **geolocation** technology dependent

Smartphones and wearables **battery duration** still remain as one of biggest handicaps.

The world extension, intermittency and stability of the **connections** are often the biggest barriers.

Mobile sensors and social media data are used, in its most simple way, as a **complement** to traditional weather observations for **verification** purposes and offline **research**. In more complex forms there are used to improve **nowcasting** and in different ways during **severe weather** episodes.

We are just **beginning** to use smartphones sensors and social media data in meteo-climatic business.

We only see the tip of the "variety" in the future internet of everything

Acknowledgements

The authors would like to thank the Emergencies and Meteorology Directorate – Security Department - Basque Government for public provision of data and operational service financial support. We also would like to thank all our colleagues from EUSKALMET for their daily effort in promoting valuable services for the Basque community

euskalmet agencia vasca de meteorología euskal meteorologia agentzia

EUSKO JAURLARITZA – SEGURTASUN SAILA – Larrialdiei Aurre Egiteko eta Meteorologiako Zuzendaritza

GOBIERNO VASCO – DEPARTAMENTO DE SEGURIDAD – Dirección de Atención de Emergencias y Meteorología

References

- Brabham, D. (2008), "Crowdsourcing as a Model for Problem Solving: An Introduction and Cases", Convergence: The International Journal of Research into New Media Technologies, 14 (1): 75–90, doi:10.1177/1354856507084420, archived from the original (PDF) on 2012-04-25
- Boyd, D.; Crawford, K. (2012). "Critical Questions for Big Data". Information, Communication & Society 15 (5): 662
- Chen H, Chiang, R.H.L. Storey V.C. 2012. Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly, 36 (4) (2012), pp. 1165–1188
- Cherkassky, D., J. Ostrometzky, and H. Messer, 2014: Precipitation classification using measurements from commercial microwave links. IEEE Trans. Geosci. Remote Sens., 52, 2350–2356, doi:10.1109/TGRS.2013.2259832.
- Doan, A.; Ramarkrishnan, R.; Halevy, A. (2011), "Crowdsourcing Systems on the World Wide Web" (ommunications of the ACM, 54 (4): 86–96, doi:10.1145/1924421.1924442
- Dotzek, N., P. Groenemeijer, B. Feuerstein, and A. M. Holzer, 2009a: Overview of ESSL's severe convective storms research using the European Severe Weather Database ESWD. Atmos. Res., 93, 575−586.
- Doumounia, A., M. Gosset, F. Cazenave, M. Kacou, and F. Zougmore, 2014: Rainfall monitoring based on microwave links from cellular telecommunication networks: First results from a West African test bed. Geophys. Res. Lett., 41, 6016–6022, doi:10.1002/2014GL060724.
- Fahrni, T. Kuhn, M. Sormmer, P. Wattenhofer, R. and Welton, S. (2011), 'Sundroid: solar radiation awareness with smartphones', UbiComp'11, Sept. 17-21, 2011, Beijing, China
- Gartner 2014. Answering Big Data's 10 Biggest Vision and Strategy Questions. Available athttps://www.gartner.com/doc/2822220?ref=SiteSearch&refval=&pcp=mpe
- Gaztelumendi, S., Egaña, J., Otxoa-de-Alda, K., Hernandez, R., Aranda, J., and Anitua, P.: An overview of a regional meteorology warning system, Adv. Sci. Res., 8, 157-166, doi:10.5194/asr-8-157-2012, 2012
- Gaztelumendi, S., Martija, M., Principe, O., and Palacio, V.: An overview of the use of Twitter in National Weather Services, Adv. Sci. Res., 12, 141-145, doi:10.5194/asr-12-141-2015, 2015.
- Gaztelumendi S. Gomez de Segura J.D. , Otxoa de Alda K. Orue J., Torres J. Aranda J.A. Big data in the context of Euskalmet activities15th EMS Annual Meeting 12th European Conference on Applications of Meteorology (ECAM) 07 – 11 September 2015, Sofia, Bulgaria 2015.
- Gill J, and Kreft P.(2012) Guidelines on Strategies for use of Social Media by National Meteorological and Hydrological Services, WMO-No. 1086, PWS-24.
- Goodchild, M.: Citizens as sensors: the world of volunteered geography, GeoJournal, 69, 211–221, 2007
- Hoffman F.M. , Larson JW Mills RT ,Bjørn-Gustaf Auroop, Ganguly, Hargrove, Huang, Kumar,Vatsavai 2011 Data Mining in Earth System Science (DMESS 2011) Procedia Computer Science. Elsevier
- Hyv¨ arinen, O. and Saltikoff, E.: Social media as a source of meteorological observations, Mon. Weather Rev., 138, 3175– 3184, 2010.
- IBM. 2015. "IBM What is big data? — Bringing big data to the enterprise". www.ibm.com.
- Igoe, Damien P. , Parisi, Alfio and Carter, Brad (2014) Smartphone based Android app for determining UVA aerosol optical depth and direct solar irradiances. Photochemistry and Photobiology, 90 (1). pp. 233-237. doi: 10.1111/php.12185 .
- Ishwarappa, J 2015 A Brief Introduction on Big Data 5Vs Characteristics and Hadoop TechnologyOriginal Research Article Pages 319- Anuradha Procedia Computer Science Volume 48, Pages 1-808
- Jan Jaap Pape, Aart Overeem, Hidde Leijnse, James Robinson Gert-Jan Steeneveld, Berthold K.P. Horn, Remko Uijlenhoe. Urban air temperature estimation from smartphone battery temperatures. 15th EMS Annual Meeting & 12th ECAM, 9 September 2015, Sofia, Bulgari
- Johnston S & James Robinson. Mobile Sensor Networks: Creating a Social Laboratory (available at http://opensignal.com/reports/2015/01/mobile-sensor-networks/)
- Kitchin R. 2014. The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. SAGE publications.
- Labrinidis A. And Jagadish H.V. 2012. Challenges and Opportunities with Big Data. Journal Proceedings of the VLDB Endowment Volume 5 Issue 12, August 2012 Pages 2032-2033
- LaValle S. Lesser, E. Shockley, R. Hopkins M.S. and Kruschwitz N. 2011. Big Data, Analytics and the Path From Insights to Value, MITSloam Management review Vol 52 No2.
- Leijnse, H., R. Uijlenhoet, and J. Stricker, 2007: Rainfall measurement using radio links from cellular communication networks. Water Resour. Res., 43, W03201, doi:10.1029/2006WR005631.
- Madden S. 2012. From Databases to Big Data IEEE Inter-net Computing, 16 (2012 June), pp. 4–6
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R. Roxburgh, C. , & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity (pp. 1–143). The McKinsey Global Institute.
- Overeem et al. (2013) Crowdsourcing urban air temperatures from smartphone battery temperatures Crowdsourcing urban air temperatures from smartphone battery temperatures GEOPHYSICAL RESEARCH LETTERS, VOL. 40, 4081–4085, doi:10.1002/grl.50786, 2013
- Overeem, A., H. Leijnse, and R. Uijlenhoet (2011), Measuring urban rainfall using microwave links from commercial cellular commu- nication networks, Water Resour. Res., 47, W12505, doi:10.1029/ 2010WR010350.
- Overeem, A., H. Leijnse, and R. Uijlenhoet (2013), Country-wide rainfall maps from cellular communication networks, PNAS, 110 , 2741–2745, doi:10.1073/pnas.1217961110.
- P. Noti, A. Martynov, A. Hering, and O. Martius. Hailstorms over Switzerland: Verification of Crowd-sourced Data. Geophysical Research Abstracts. Vol. 18, EGU2016-5822, EGU General Assembly 2016.
- Schroeck M, Shockley R, Smart J, Romero-Morales D , Tufano P 2012. Analytics: The real-world use of big data Business Analytics and Optimization. IBM Institute for Business Value Report.
- Singh D. and Reddy C.K. 2014 "A survey on platforms for big data analytics", Journal of Big Data, Vol.2, No.8, pp.1-20, October 2014.
- Snik, F., et al. (2014), Mapping atmospheric aerosols with a citizen science network of smartphone spectropolarimeters, Geophys. Res. Lett., 41, 7351–7358, doi:10.1002/2014GL061462.
- V. Parisi, N. Downs, D. Igoe, J. Turner, Characterization of cloud cover with a smartphone camera, Instrumentation Science & Technology, 2016, 44, 1, 23

Contact info:
santiago.gaztelumendi@tecnalia.com
josedaniel.gomezdesegura@tecnalia.com

TECNALIA
Parque Tecnológico de Bizkaia
C/ Geldo Edificio 700
E-48160 DERIO (Bizkaia) Spain
www.tecnalia.com

16th EMS Annual Meeting
11th European Conference on Applied Climatology (ECAC)
12–16 September 2016 , Trieste, Italy