CONSORZIO INTERUNIVERSITARIO PER L'IDROLOGIA

United Nations Educational, Scientific and Cultural Organization · World Water Assessment Programme

WA RRE DOC

Università per Stranieri di Perugia

2019 Edition
Doctoral Winter School
DATA RICH HYDROLOGY

Citizen Science and Big Data In Hydrology

Part 1. Geomorphic signatures of hydrologic processes using global topographic datasets

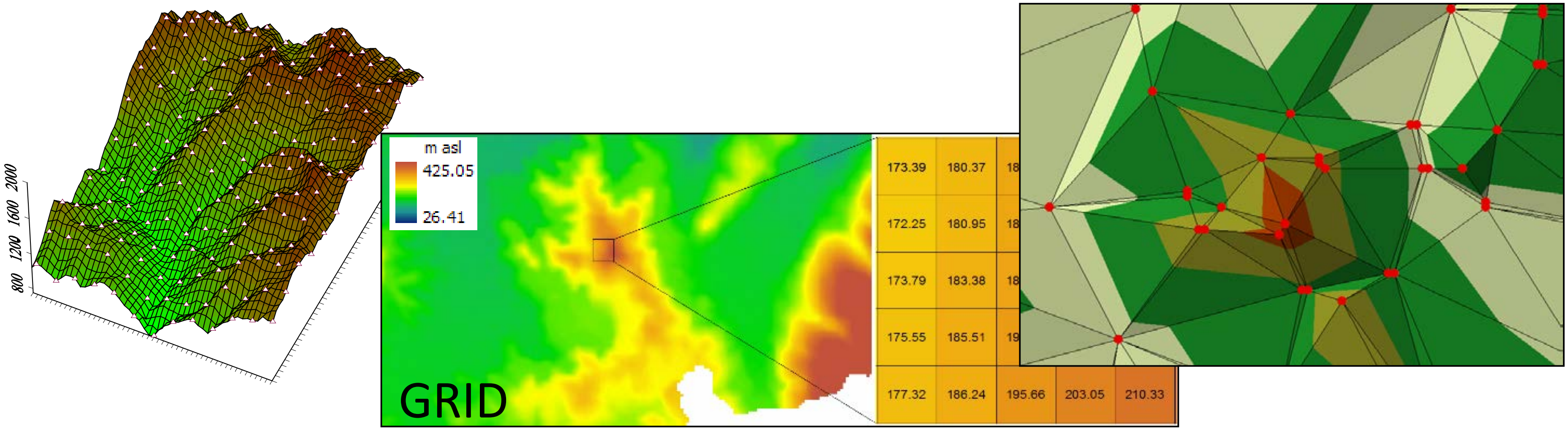
January 31, 2019

Fernando Nardi

Summary

- DEMs and Terrain Analysis using global topographic datasets
- Hydrologic features (rivers, floodplains, …) identification
- Open/Big data and digital geography

Digital Topography



DTMs are GIS layers representing absolute terrain elevations in **raster (grid/matrix)** and **TIN (Triangular Irregular Network)** format

DEM attributes

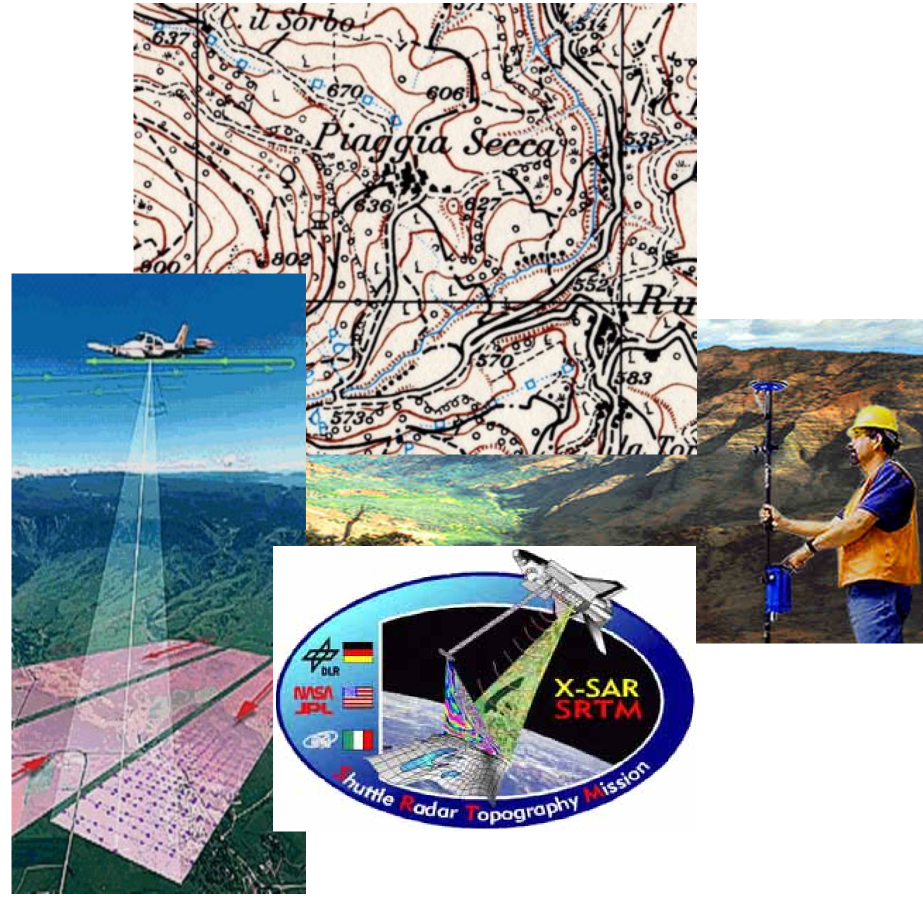
Resolution: Spacing of the elevation points
Precision: Floating/Integer values
Accuracy: Spatial and vertical RMSE

DTM data sources

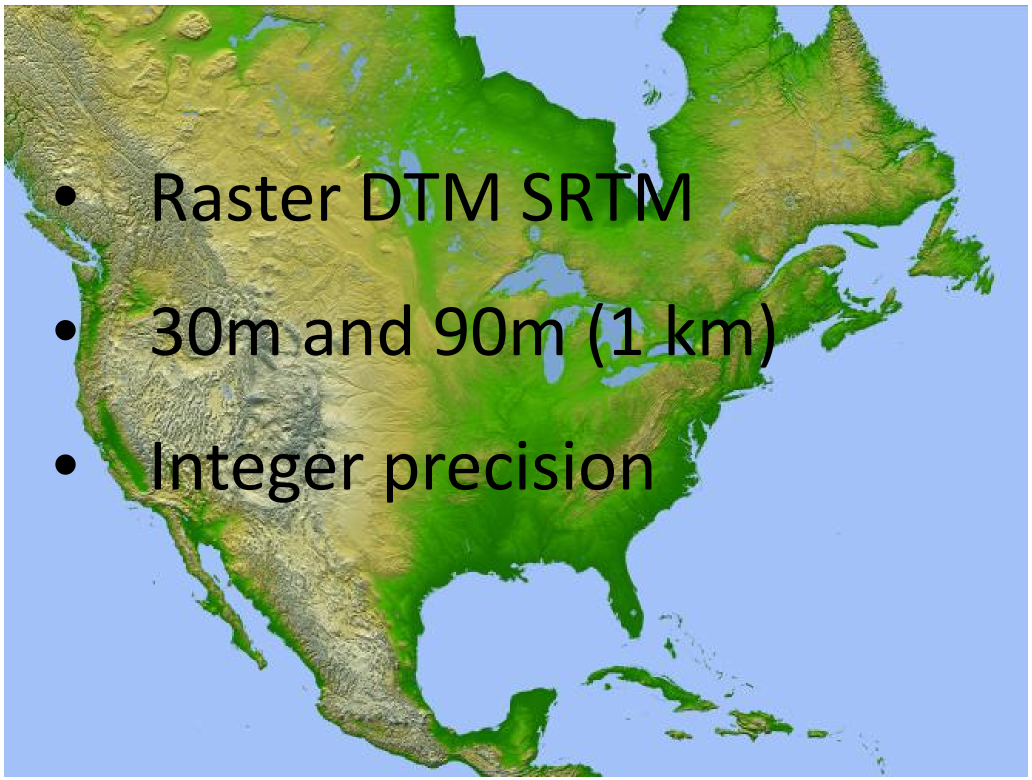
Topographic maps

Field surveys
(GPS, Laser scanner)

Aerial and satellite remete sensing
(Lidar, Drones, SRTM, ASTER)



SRTM



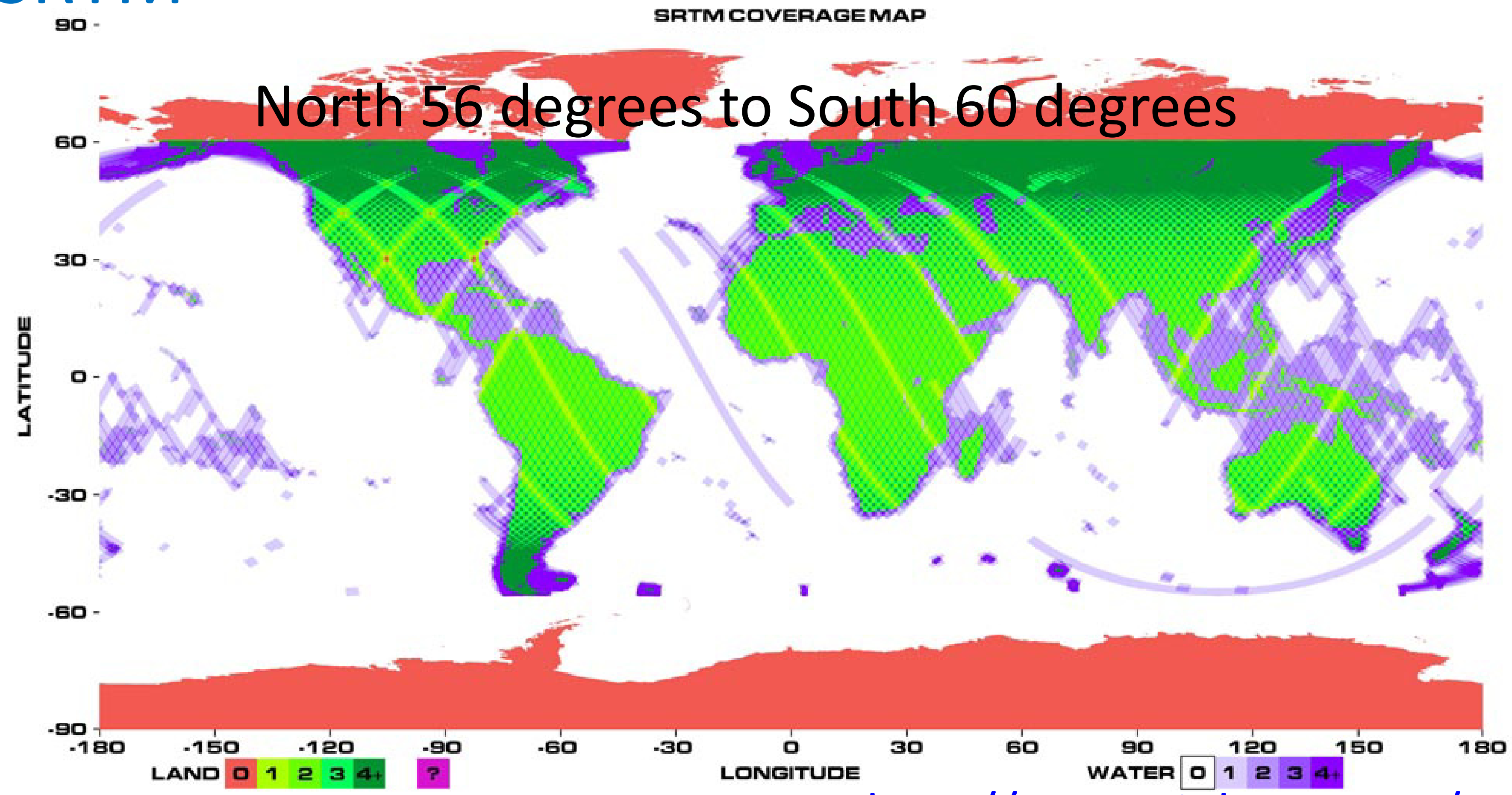
- Raster DTM SRTM
- 30m and 90m (1 km)
- Integer precision

Target Name:	Earth
Is a satellite of:	Sol (our sun)
Mission:	Shuttle Radar Topography Mission (SRTM)
Spacecraft:	Space Shuttle Endeavour
Instrument:	C-Band Interferometric Radar
Product Size:	9600 x 7240 pixels (width x height)
Produced By:	JPL
Full-Res TIFF:	PIA03377.tif (130.7 MB)
Full-Res JPEG:	PIA03377.jpg (56.09 MB)

http://www2.jpl.nasa.gov/srtm/

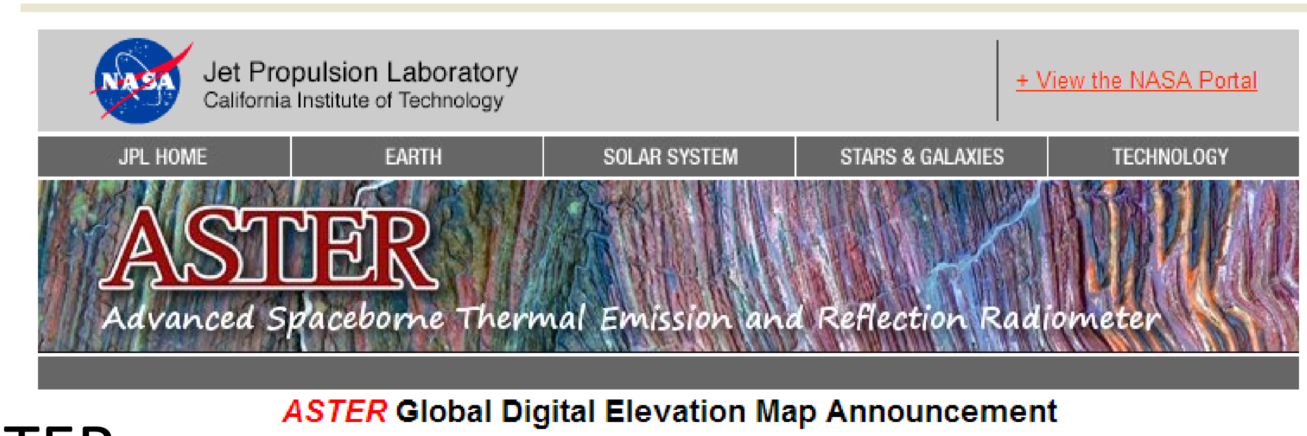
SRTM

North 56 degrees to South 60 degrees



http://www2.jpl.nasa.gov/srtm/

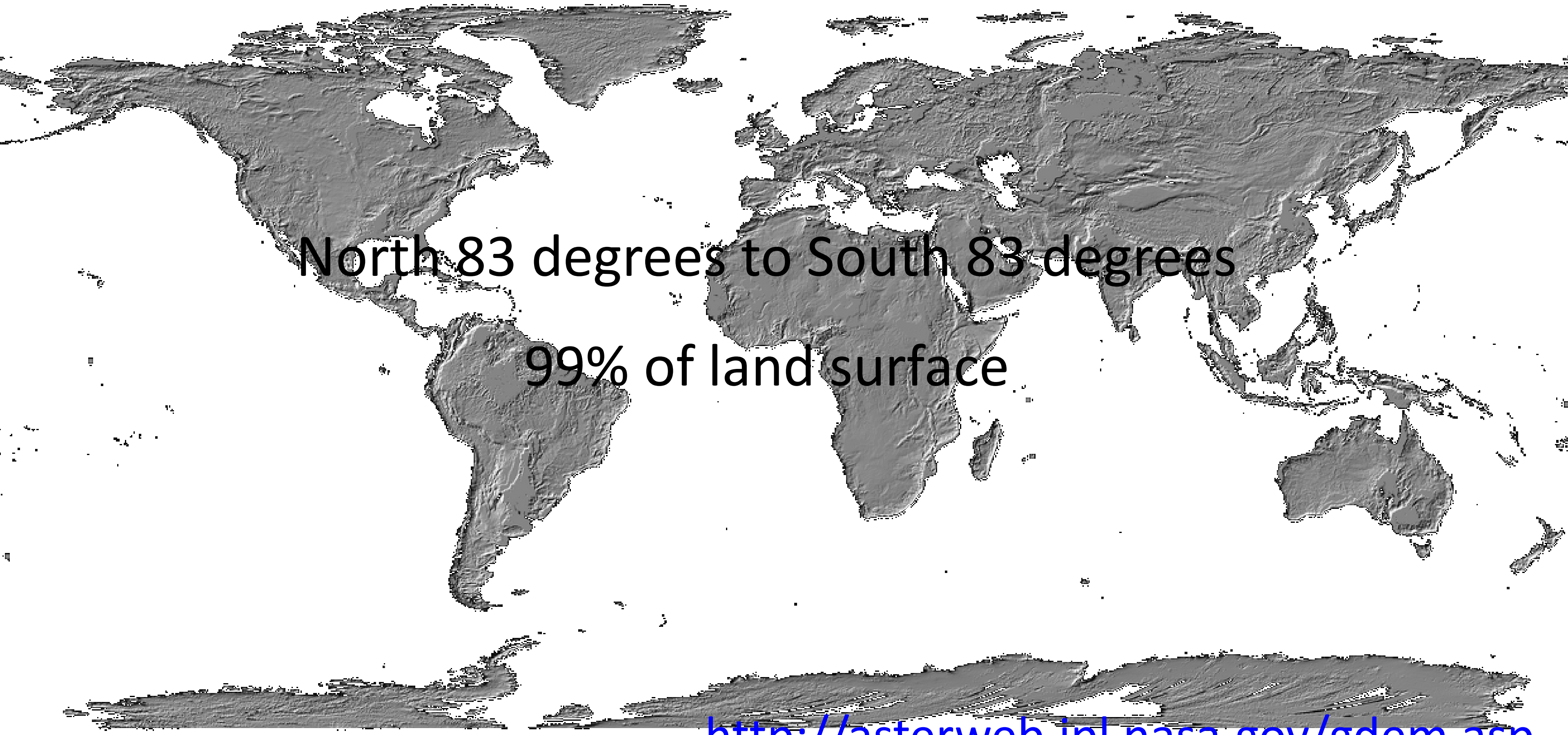
ASTER



- Raster DTM ASTER
- 30m
- Integer precision

http://asterweb.jpl.nasa.gov/gdem.asp

ASTER



North 83 degrees to South 83 degrees

99% of land surface

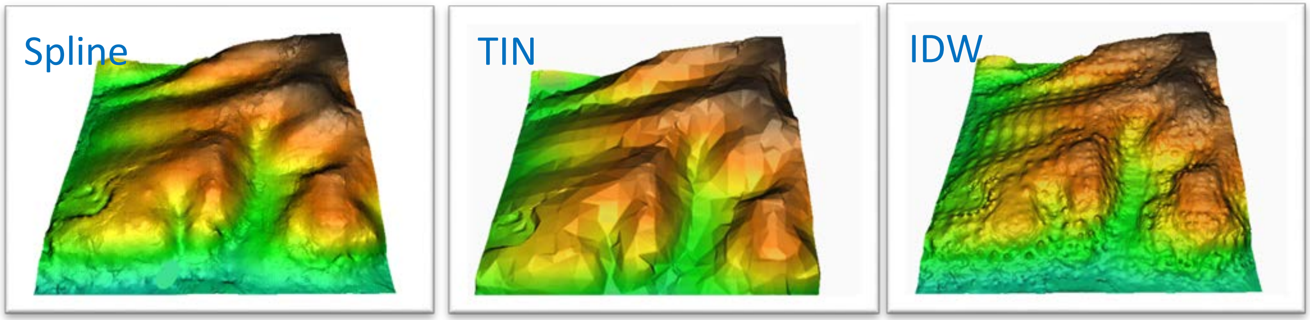
http://asterweb.jpl.nasa.gov/gdem.asp

Commercial products

http://www.geoeye.com
http://apollomapping.com/
http://it.wikipedia.org/wiki/Immagini_satellitari

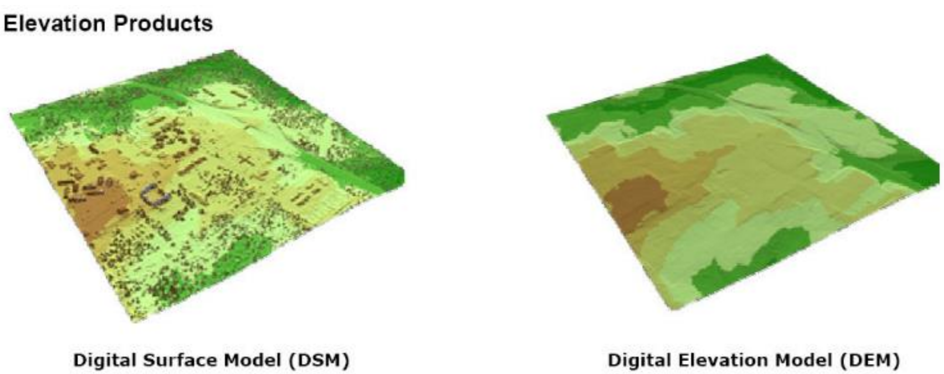
Major factors affecting the DEM

Interpolation techniques (data voids)



From
Graeme Aggett 2001

Postprocessing for interpreting vegetation and urban features



Digital Earth

PNAS PNAS

Next-generation Digital Earth

PERSPEC

Michael F. Goodchild[a,1], Huadong Guo[b], Alessandro Annoni[c], Ling Bian[d], Kees de Bie[e], Frederick Campbell[f], Max Craglia[c], Manfred Ehlers[g], John van Genderen[e], Dav
Richard Simpson[k], Andrew Skidmore[f], Cha
[a]Department of Geography, University of Califo
Academy of Sciences, Beijing 100094, China; [c]Joi
Geography, University at Buffalo, State Universit
Observation, University of Twente, 7500 AE, Ens
[g]Institute for Geoinformatics and Remote Sensin
Australia; [i]Department of Geography and Anthr
Geo-Information, H-1122, Budapest, Hungary; [k]N
Information, Carlton South 3053, Australia

Edited by Kenneth Wachter, University of California, B

A speech of then-Vice President Al Gore in 1998 cr
generation of virtual globes, typified by Google E
of Google Earth, and the functionality of this fi
developments in technology continue, the era o
through citizen science and crowd-sourcing, and
although Google Earth stimulated progress in c
public's access to science. All these factors prom
elements that should be part of a next generati

Google

Search Places or Keywords...

Earth Engine

A planetary-scale platform for environmental data & analysis

Google Earth Engine brings together the world's satellite imagery — trillions of scientific measurements dating back almost 40 years — and makes it available online with tools for scientists, independent researchers, and nations to mine this massive warehouse of data to detect changes, map trends and quantify differences on the Earth's surface. Applications include: detecting deforestation, classifying land cover, estimating forest biomass and carbon, and mapping the world's roadless areas.

To learn more, view product videos and the Featured Gallery (below). Or visit the Data Catalog to explore our archive of satellite imagery. Certain features (such as data download) are restricted to members of our trusted tester program.

New! Congratulations to the Landsat program on 40 years of continuous Earth observation! Learn more...

Earth Engine API

Develop, access and run algorithms on the full Earth Engine data archive, all using Google's parallel processing platform.

The Earth Engine API is currently available as a limited release to a small group of partners. If you are interested in developing on the Earth Engine platform, let us know.

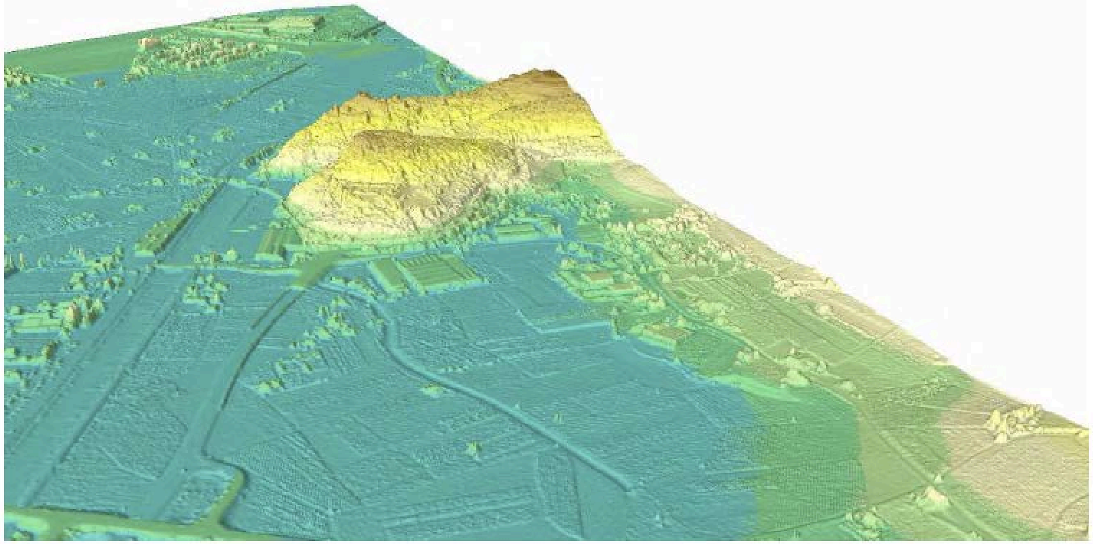
Featured gallery

We have precomputed a number of interesting datasets using the Earth Engine platform, below. Click on each to learn more and preview the data as a global time-lapse or as a layer in a Google Earth client.

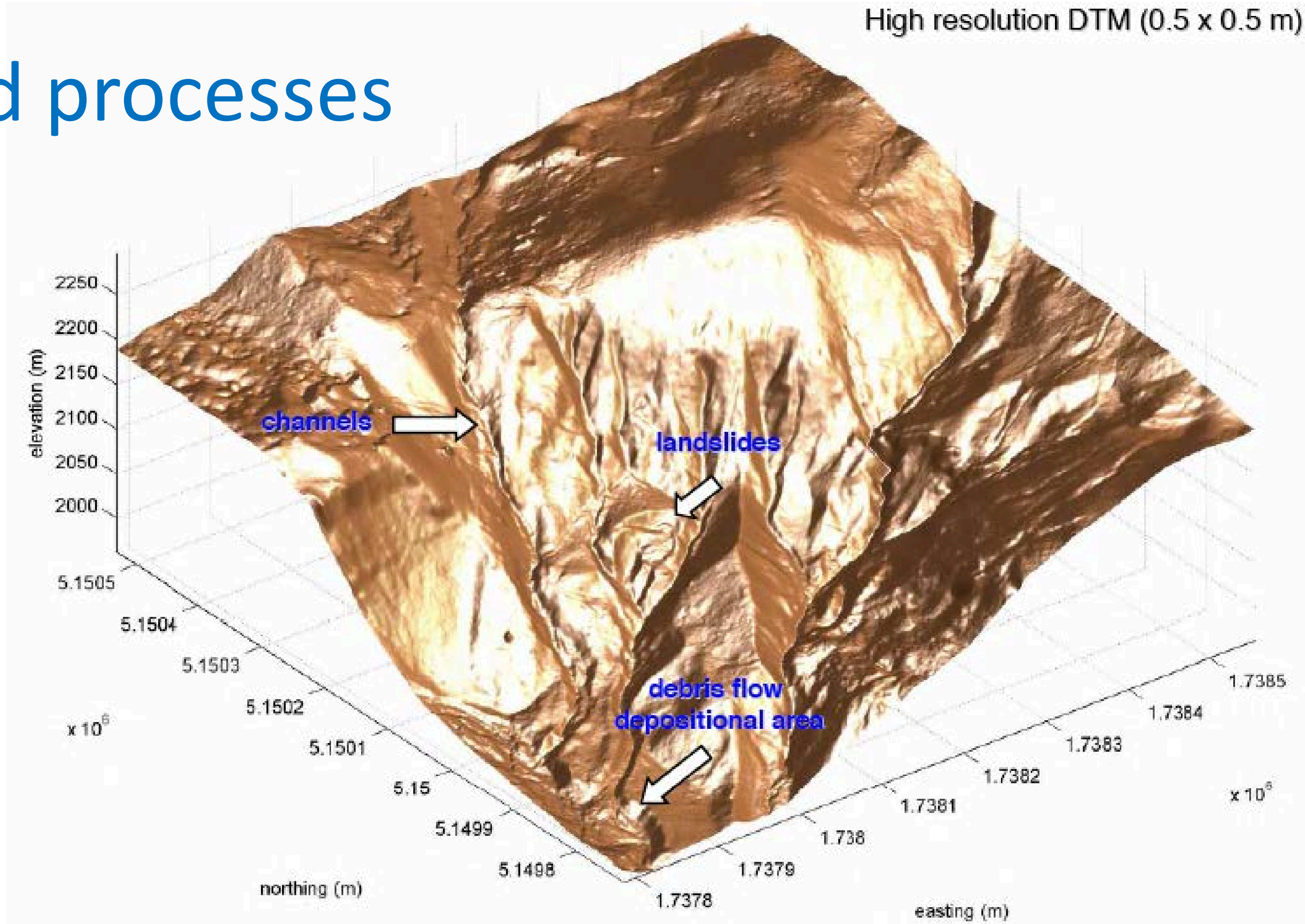
Growth of Las Vegas: Timelapse
Interactive Landsat timelapse of urban expansion and water resources in the Nevada desert.

Amazon Deforestation: Timelapse
Interactive Landsat timelapse of deforestation of the Amazon rainforest, 1999-2011.

Observing urban and natural features…



…and processes



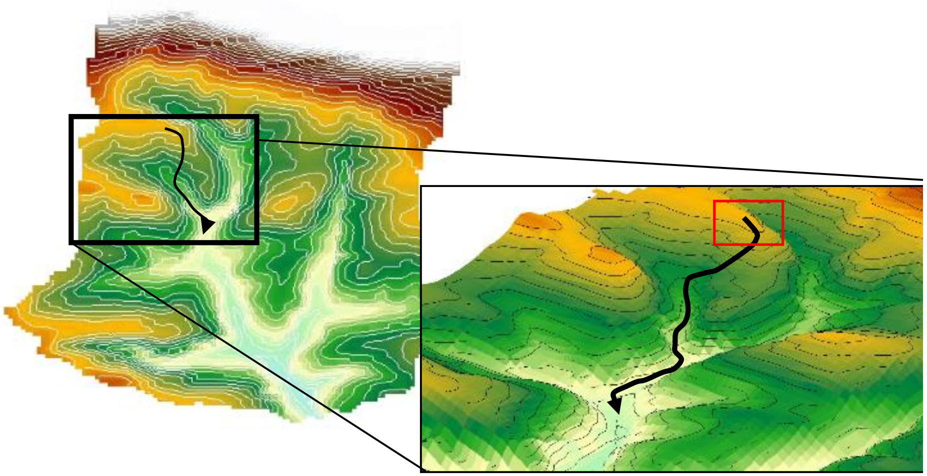
Passalacqua et al. (2010). Testing space-scale methodologies for automatic geomorphic feature extraction from lidar in a complex mountainous landscape. Water resources research, 46(11).

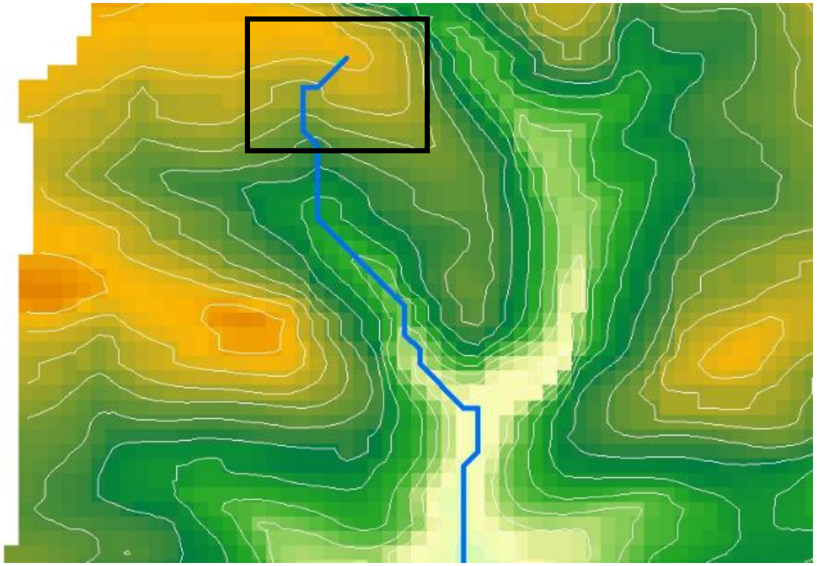
Analyzing terrain surface properties (elevation, land use,..) for deriving parameters:

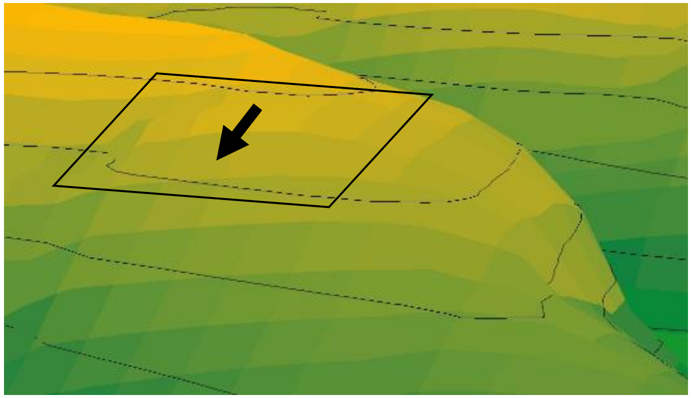
Direct: slope, curvature, aspect
Indirect: flow direction, drainage area

Features:
stream network, drainage basin, floodplain, hillslope

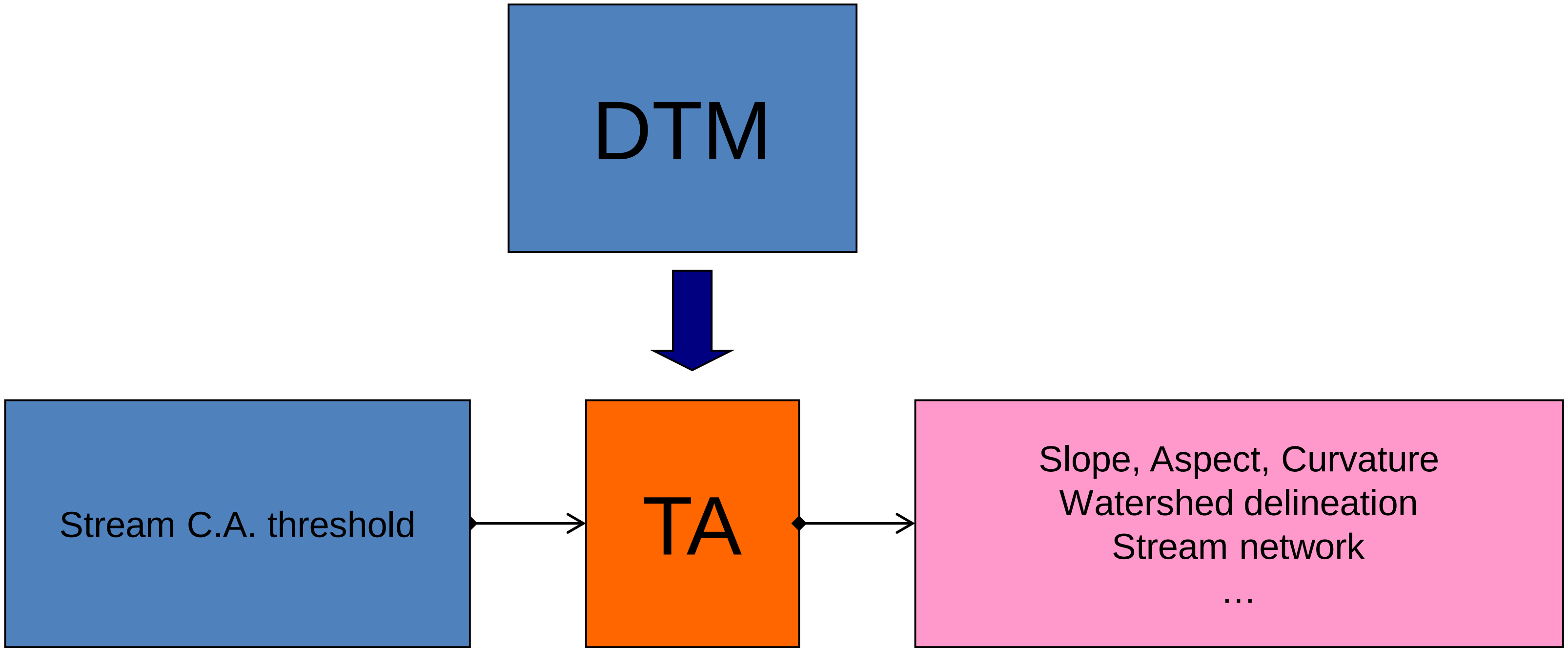
Processes:
deposition, erosion, land/water dynamics (i.e. landsliding/flooding)



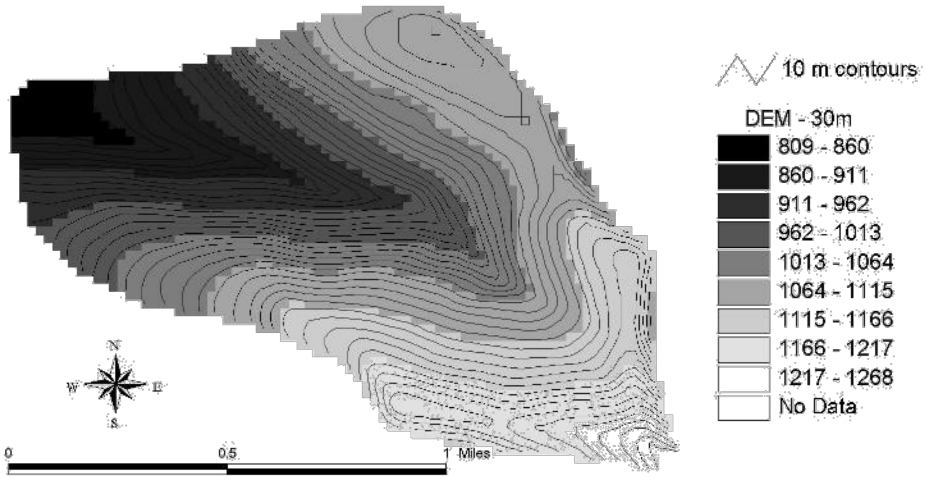


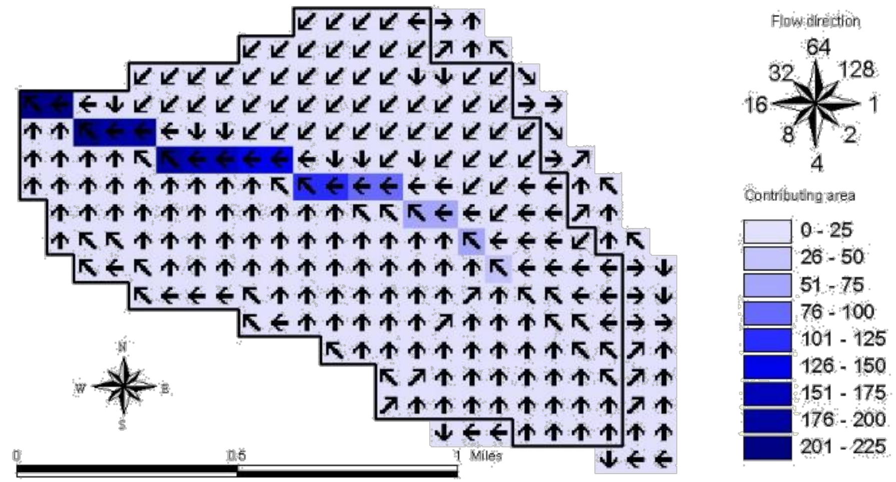


Water flows along the steepest slope



Stream network extraction





Filling sinks
Flow direction
Contributing area
Watershed delineation
Stream Network

Hydrologic processing of DTMs for stream network extraction

Tarboton, D. G., Bras, R. L., & Rodriguez-Iturbe, I. (1991). On the extraction of channel networks from digital elevation data. *Hydrological processes*, *5*(1), 81-100.

Floodplains

- Riverine valleys are unique landscape features that have supported human development since antiquity
- Fluvial or riverine corridors, stretching from mountains to coastal areas, process gravity-controlled fluxes of water, sediments and nutrients giving shape to human, plant and animal dynamics
- Floodplains are among the most resource valuable and life rich environments as well as the most endangered and risk-prone areas on earth



Global floodplain model

- Global flood hazard models generally require physically based Hydrologic and Hydraulic (H&H) modelling for simulating flood wave propagation
- River basin landscape morphology (natural and urban feature geometry -> topography) is a governing factor of flood hazard models
- Computational burden and urban data resolution/accuracy impact global flood hazard models
- Event-based vs continuous (real time) hydrologic modelling/calibration for design hydrograph / flood wave analysis is a highly uncertain condition (climate/hydrologic change)

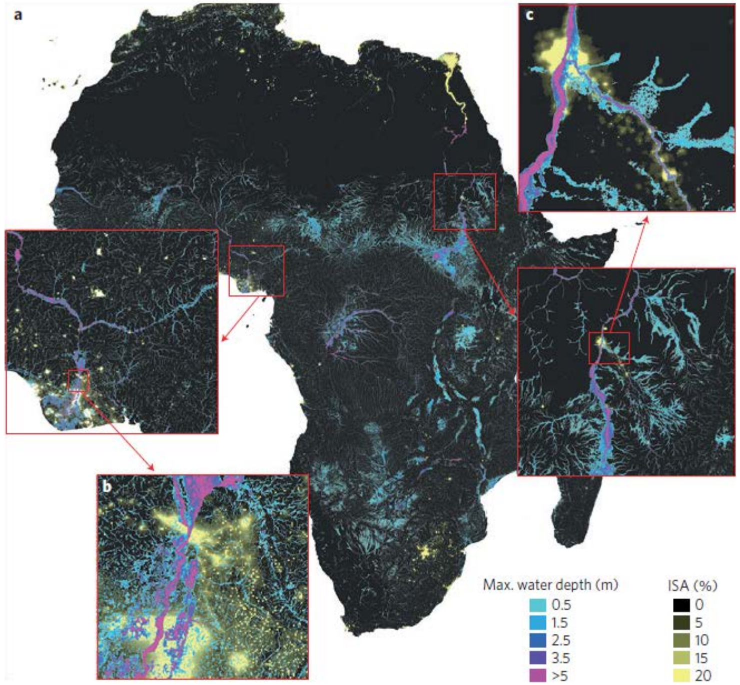


Figure 1 | Example hazard data from a global flood model, overlaid on impervious surface area (ISA) data[27] as an indicator of exposure. The hazard data shown here are from the SSBN global flood model[14], and show 1-in-100-year maximum flood depth for: **a**, all of Africa; **b**, River Niger at Onitsha, Nigeria; and **c**, Blue Nile at Omdurman, Sudan.

Ward et al. (Nature Climate Change, 2015)

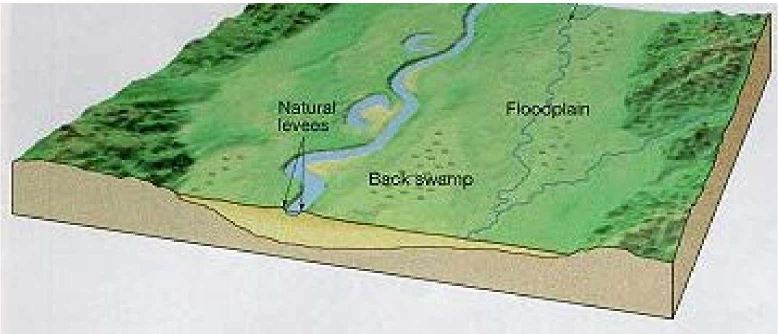
Floodplains

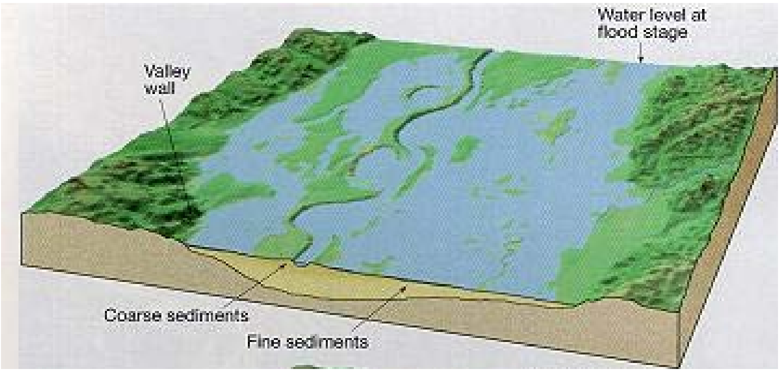
Floodplains are critical landscape features that are highly distinguished from neighboring uplands in terms of their hydrologic, geomorphologic, biogeochemical and ecological processes

Nardi, F., E. R. Vivoni, and S. Grimaldi (2006), Investigating a floodplain scaling relation using a hydrogeomorphic delineation method, Water Resources Research, 42

Floodplains

Large floods leave a **clear trace on fluvial valleys**, leading to distinct hydrogeomorphic, biogeochemical and ecologic properties in floodplains with respect to surrounding regions.



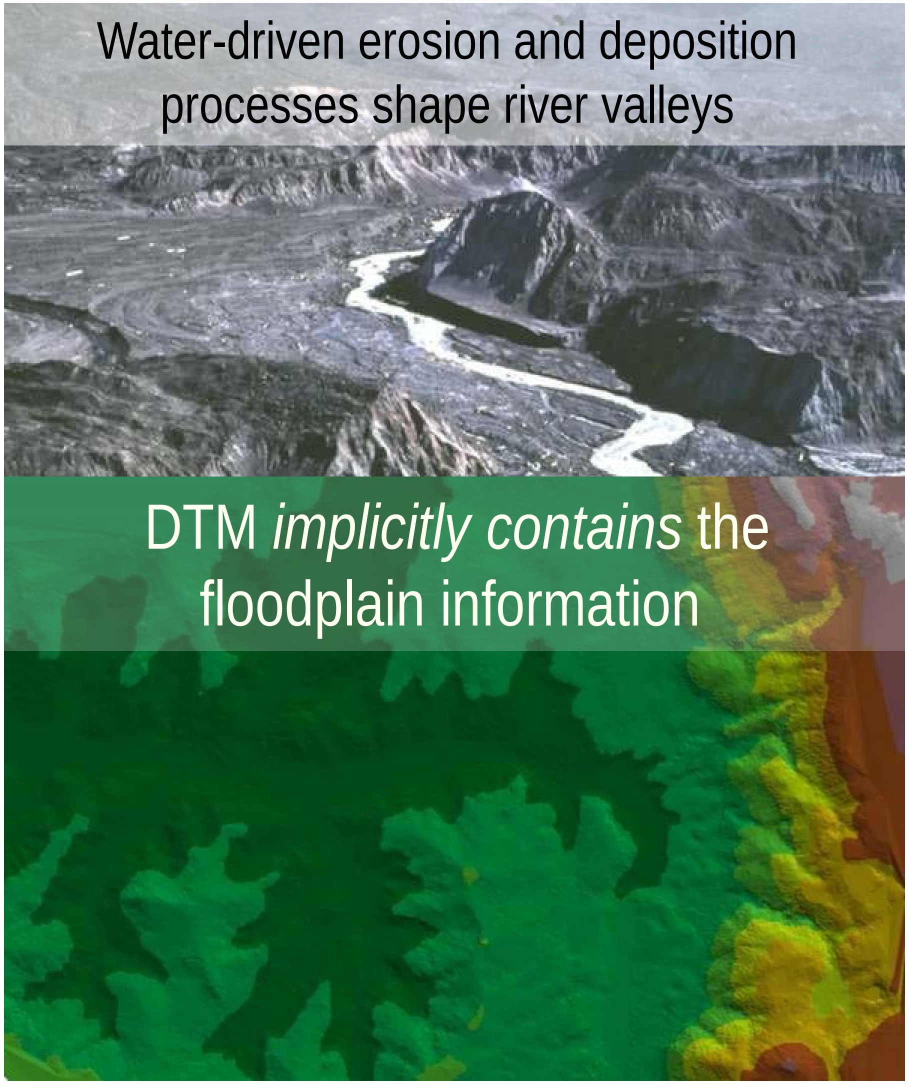


Nardi, F., E. R. Vivoni, and S. Grimaldi (2006), Investigating a floodplain scaling relation using a hydrogeomorphic delineation method, Water Resources Research, 42

Floodplains

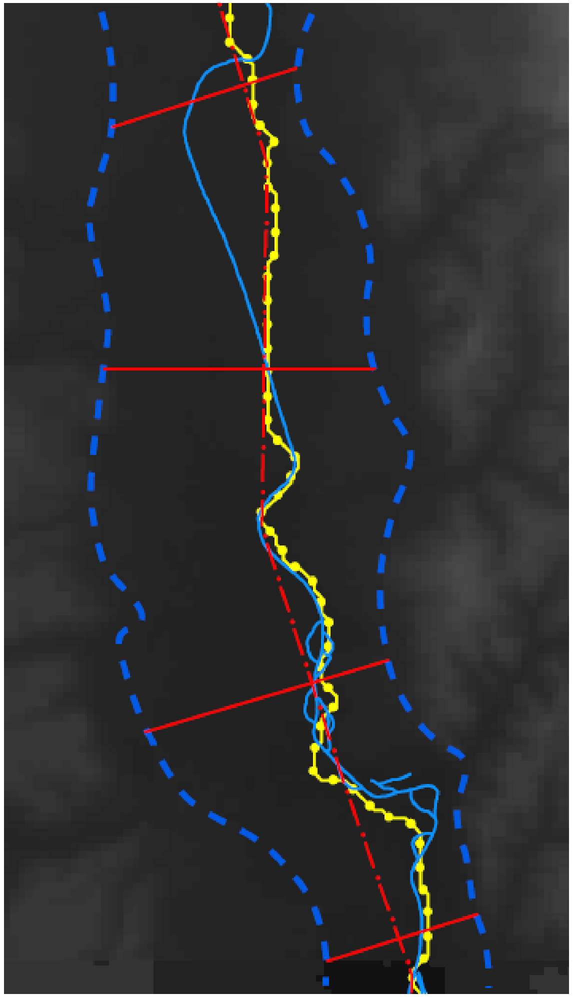
Floodplain topography intrinsically contains information on the hydrogeomorphic signature of past erosional and depositional processes.

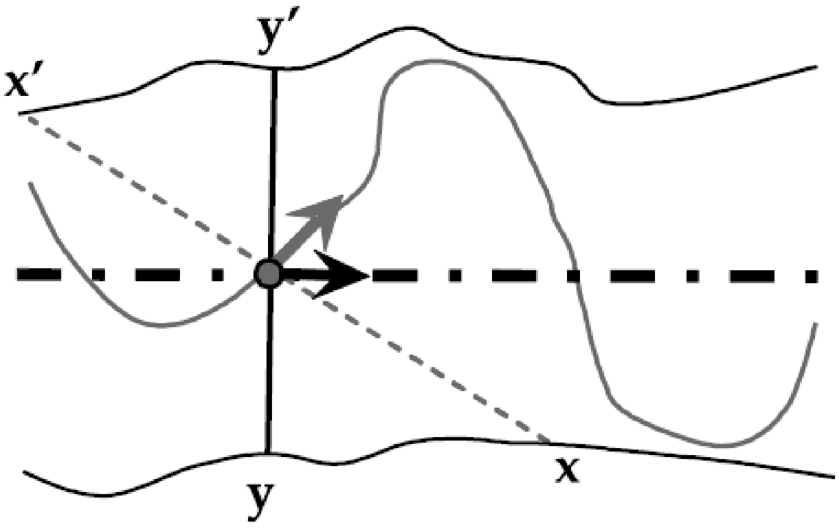
Digital Terrain Models (DTMs) contains significant information on the floodplain-related physical phenomena

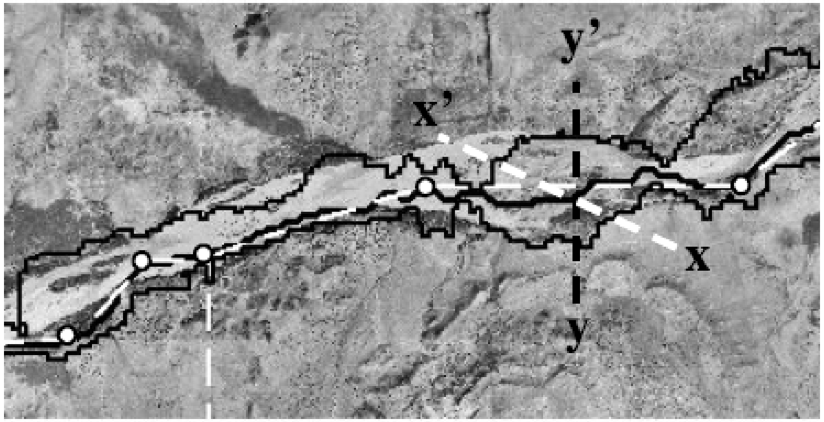


Nardi, F., E. R. Vivoni, and S. Grimaldi (2006), Investigating a floodplain scaling relation using a hydrogeomorphic delineation method, Water Resources Research, 42

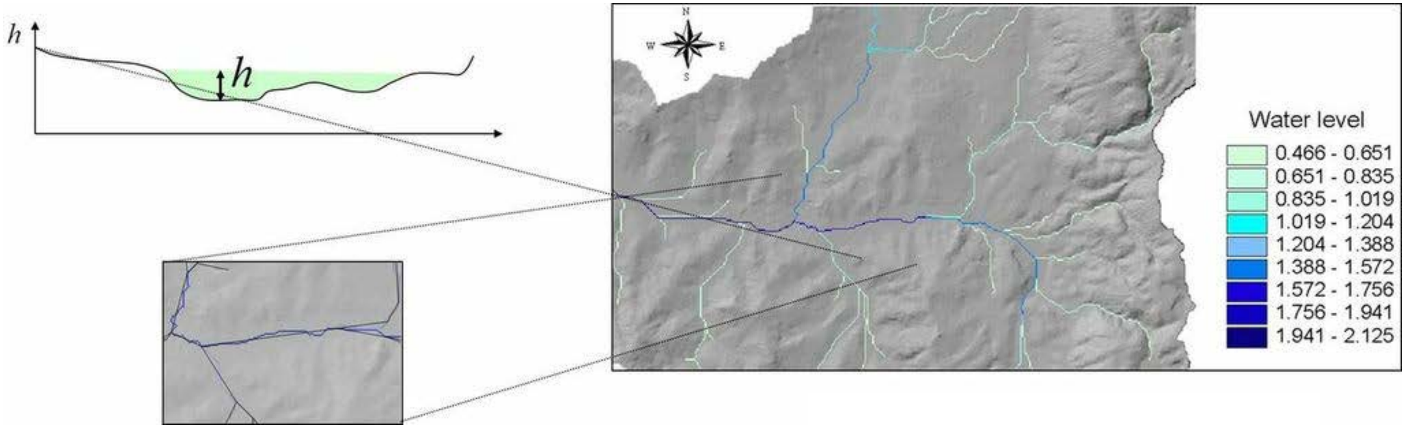
Floodplain cross section identification

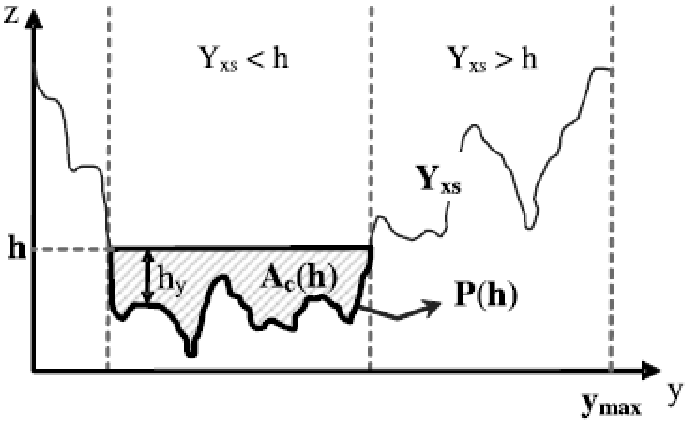






Inundation depth



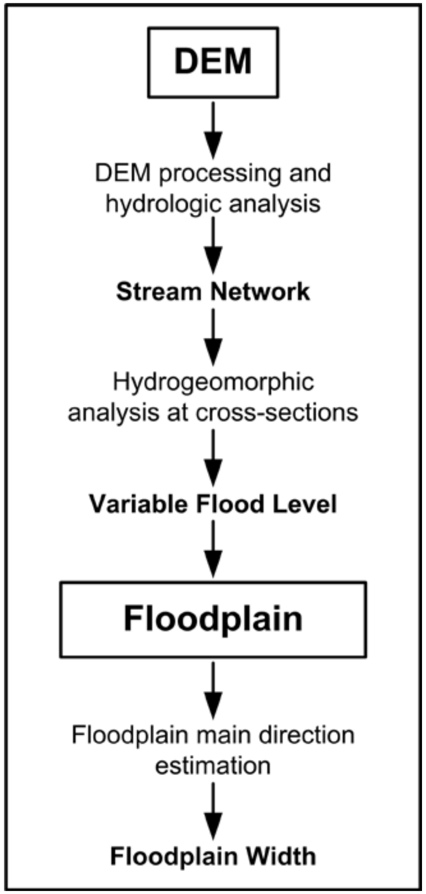


$$Q_o = \left[\frac{1.31}{L_\Omega} R_l^{0.43} A_b\right] vi$$

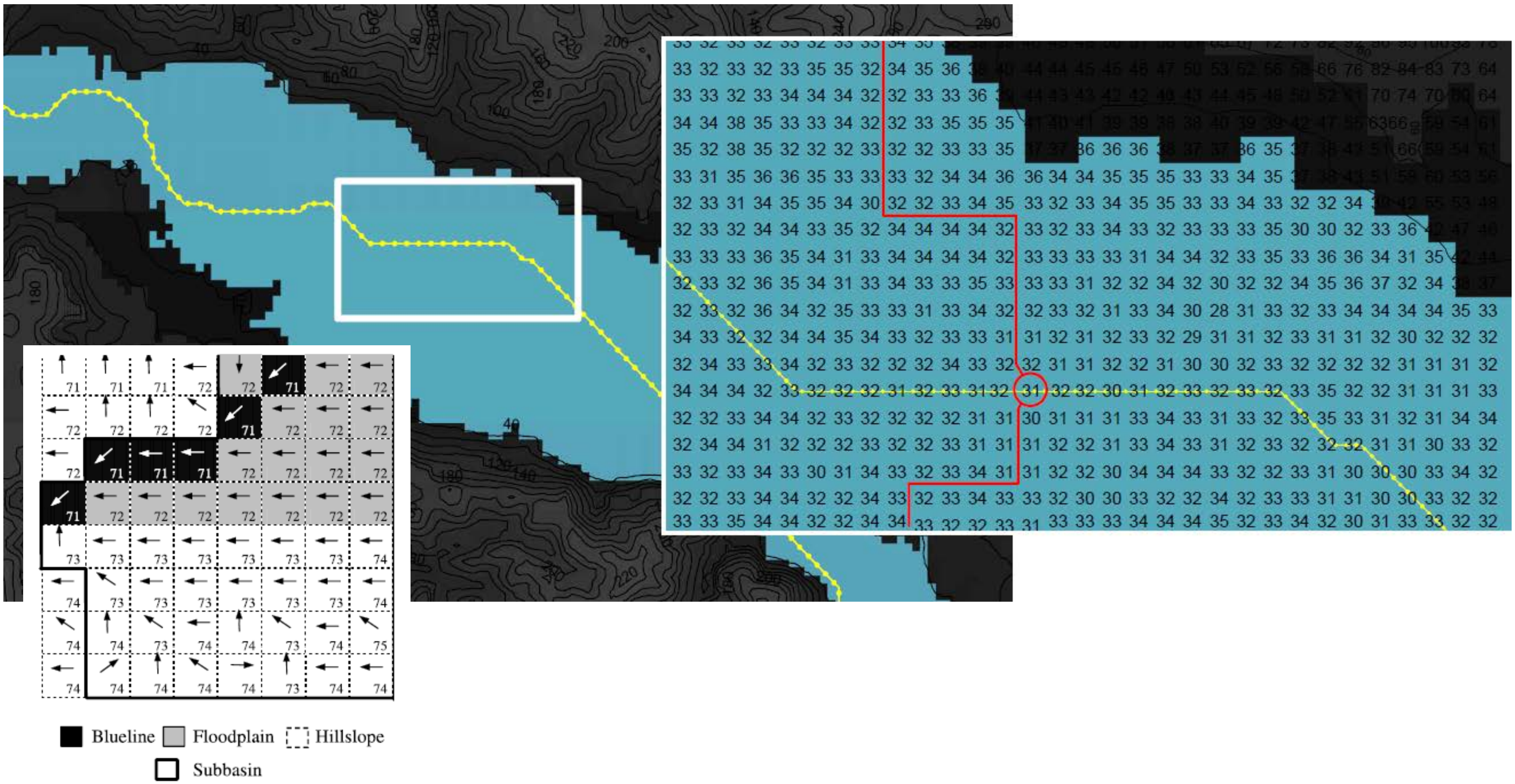
River network geomorphic properties
Climatic conditions (Rainfall)

$$Q_c = \frac{1}{n} A_c R^{1/3} S^{1/2}$$

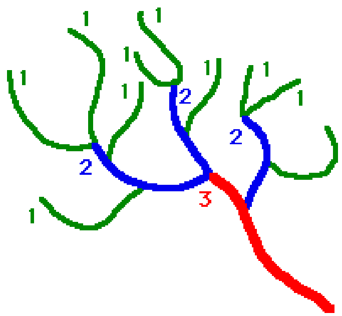
Local hydrogeomorphic properties
Of the floodplain (fluvial) cross section

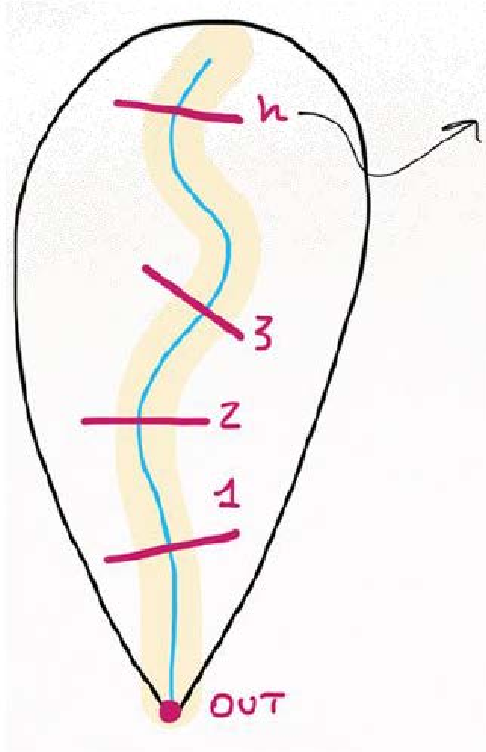


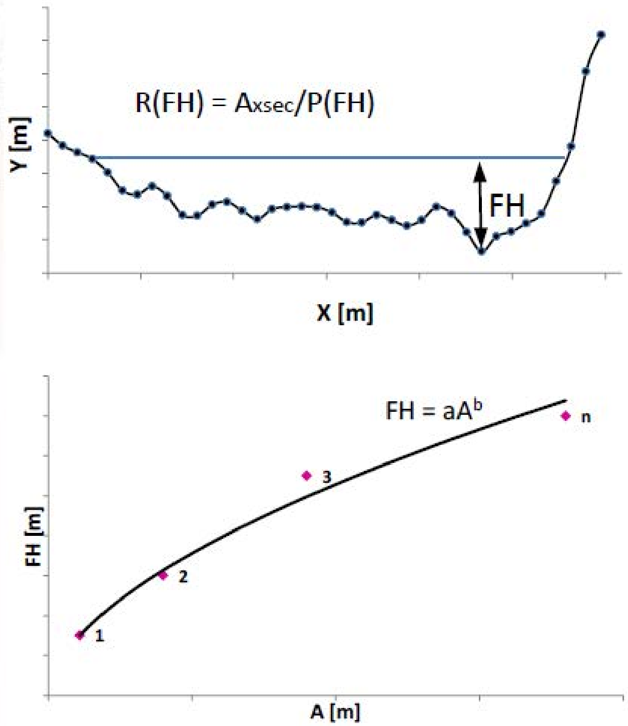
Flood prone areas delineation



Leopold law parameters estimation

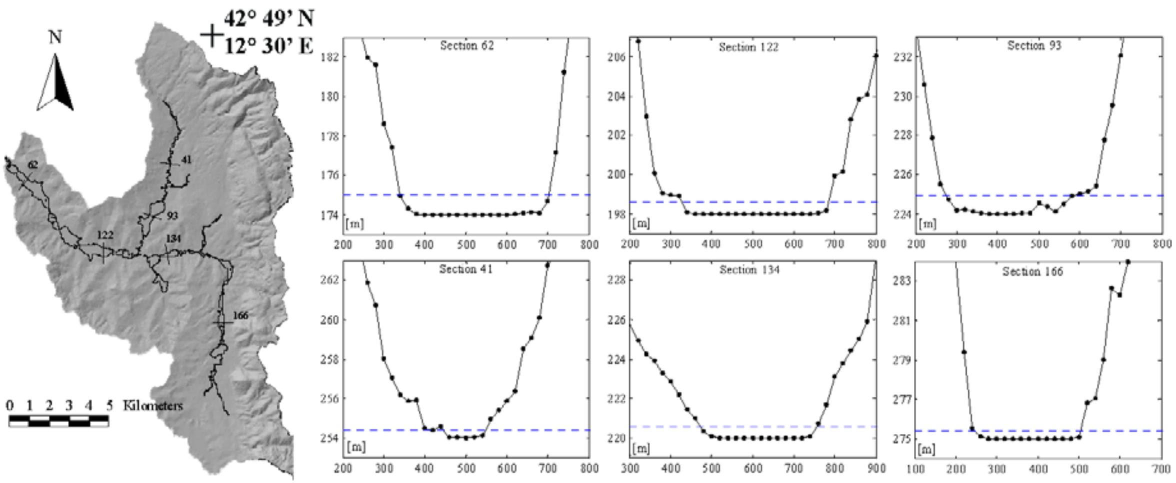


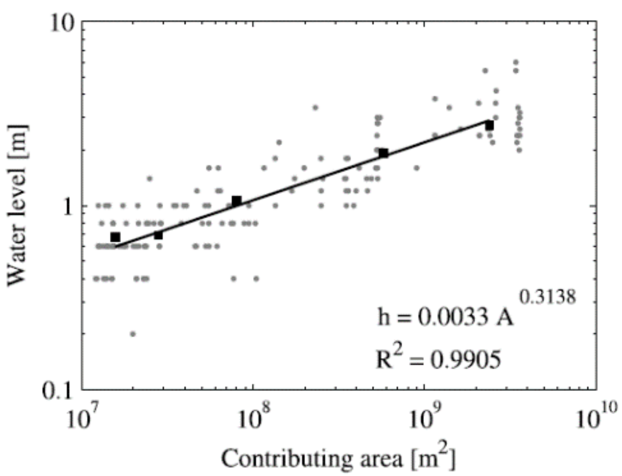




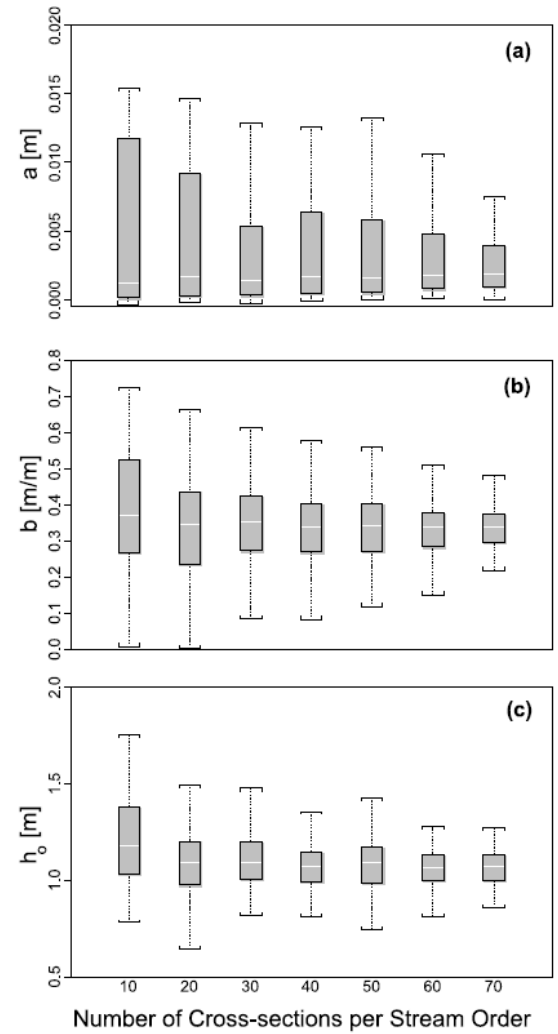
Inundation depths are averaged per stream order and the scaling parameters of Leopold's law are estimated

Leopold law parameters estimation

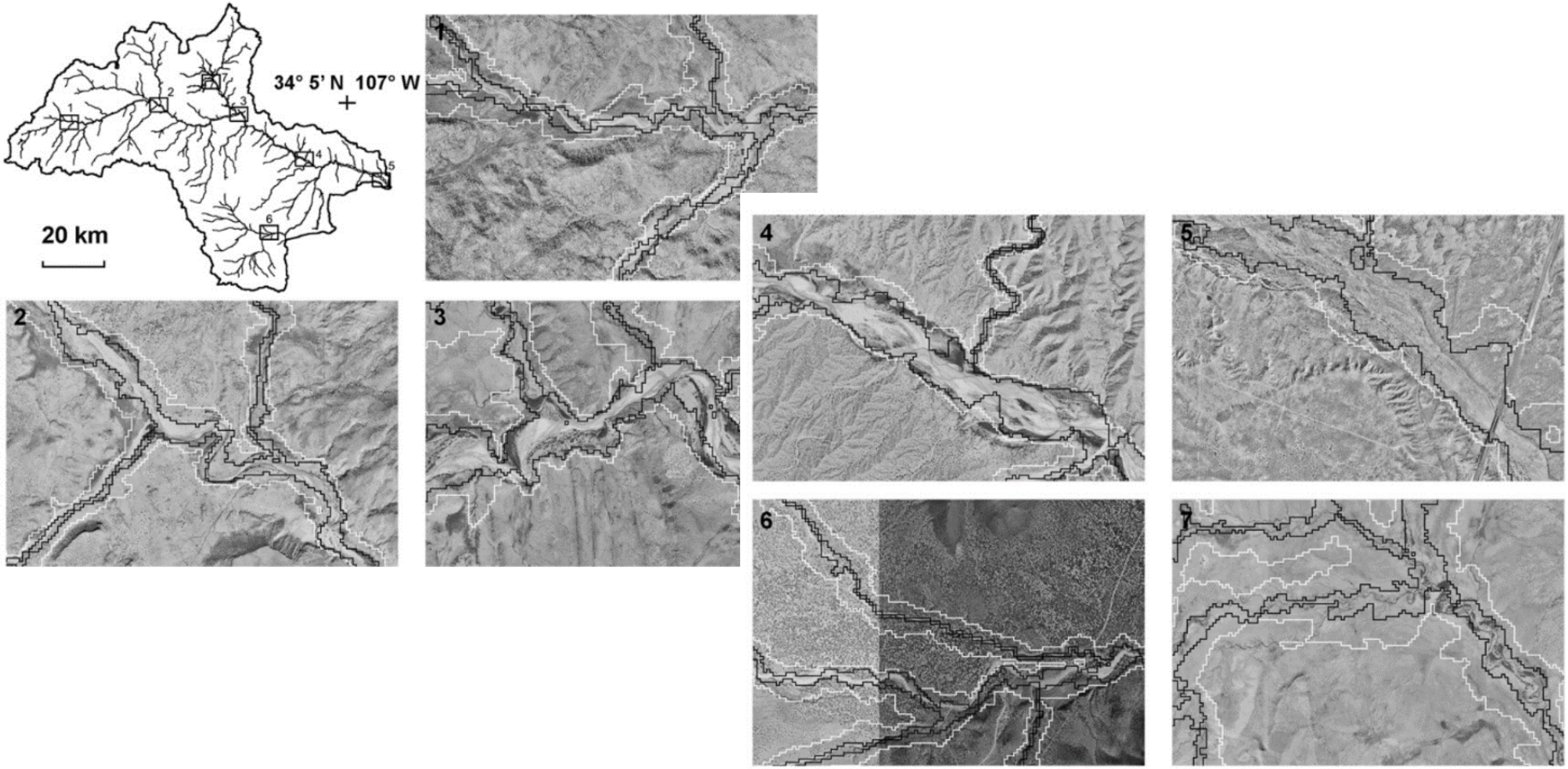




Manning Roughness n	Intercept a, m	Slope b, m/m
0.03	0.0032 (0.0010)	0.311 (0.017)
0.04	0.0033 (0.0009)	0.315 (0.013)
0.05	0.0034 (0.0009)	0.319 (0.014)
0.07	0.0035 (0.0006)	0.322 (0.010)



Results



Floodplain width scaling

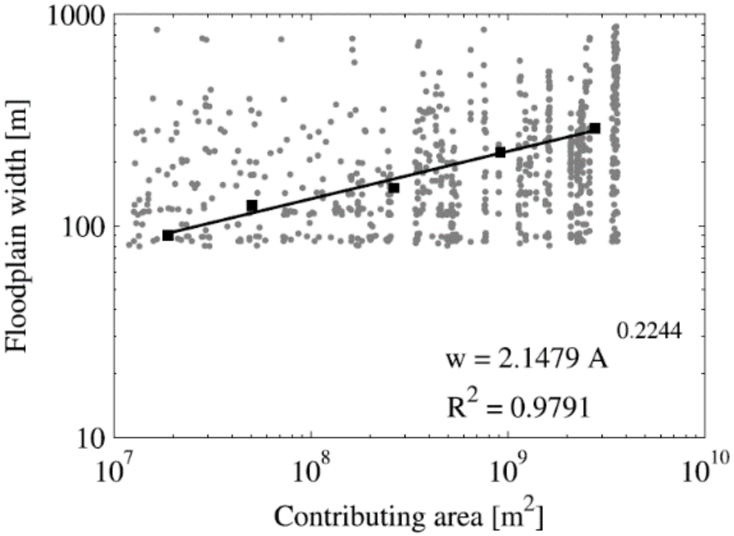
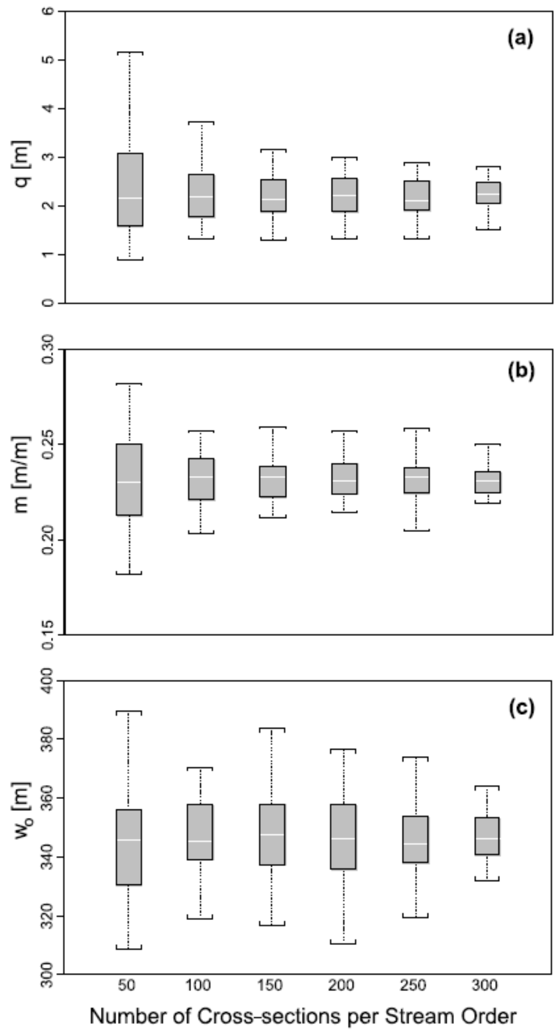
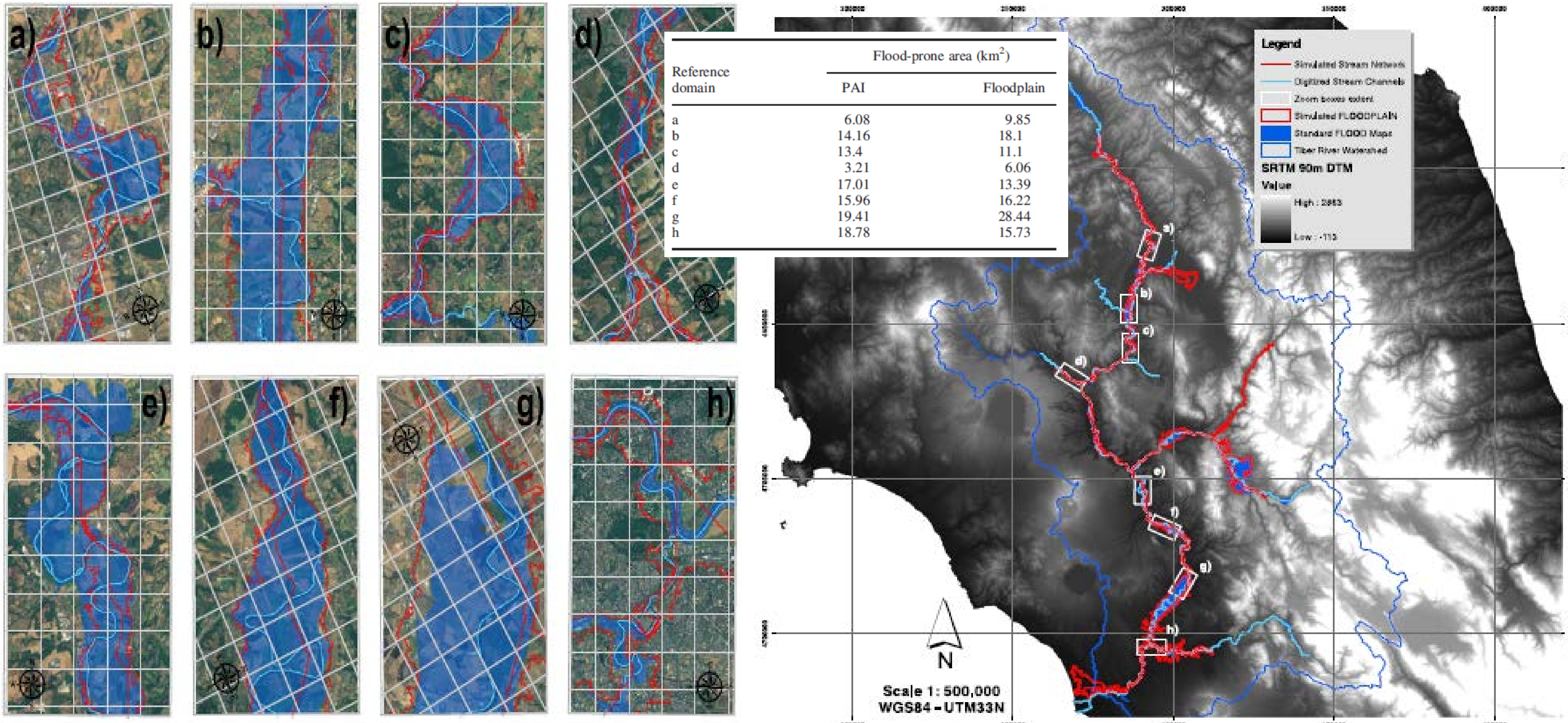


Table 5. Floodplain Width Scaling Parameters (q, m) for the Río Salado Under Varying Return Periods (T_r)[a]

Return Period T_r, years	Intercept q, m	Slope m, m/m
10	4.43 (0.79)	0.180 (0.009)
50	3.04 (0.48)	0.202 (0.008)
100	2.52 (0.45)	0.213 (0.009)
200	2.42 (0.35)	0.225 (0.009)
500	2.24 (0.33)	0.231 (0.008)



How do the hydrogeomorphic floodplain behave as respect to standard flood/inundation maps?

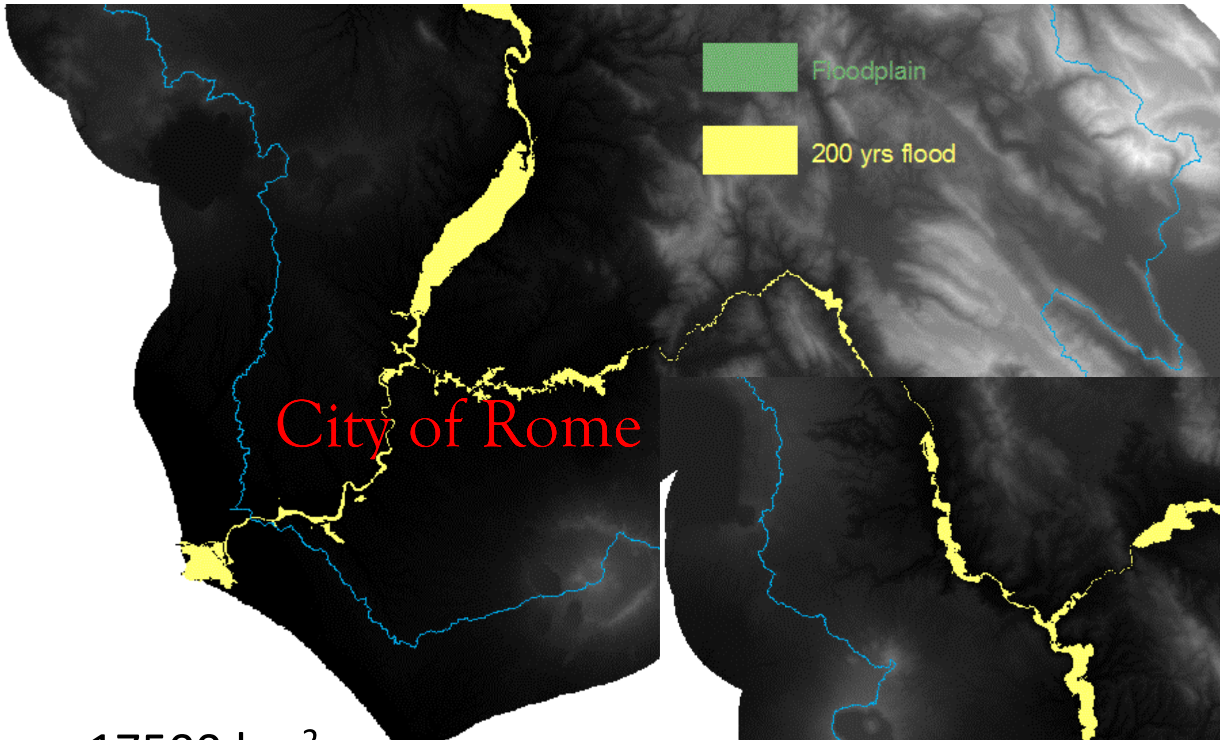


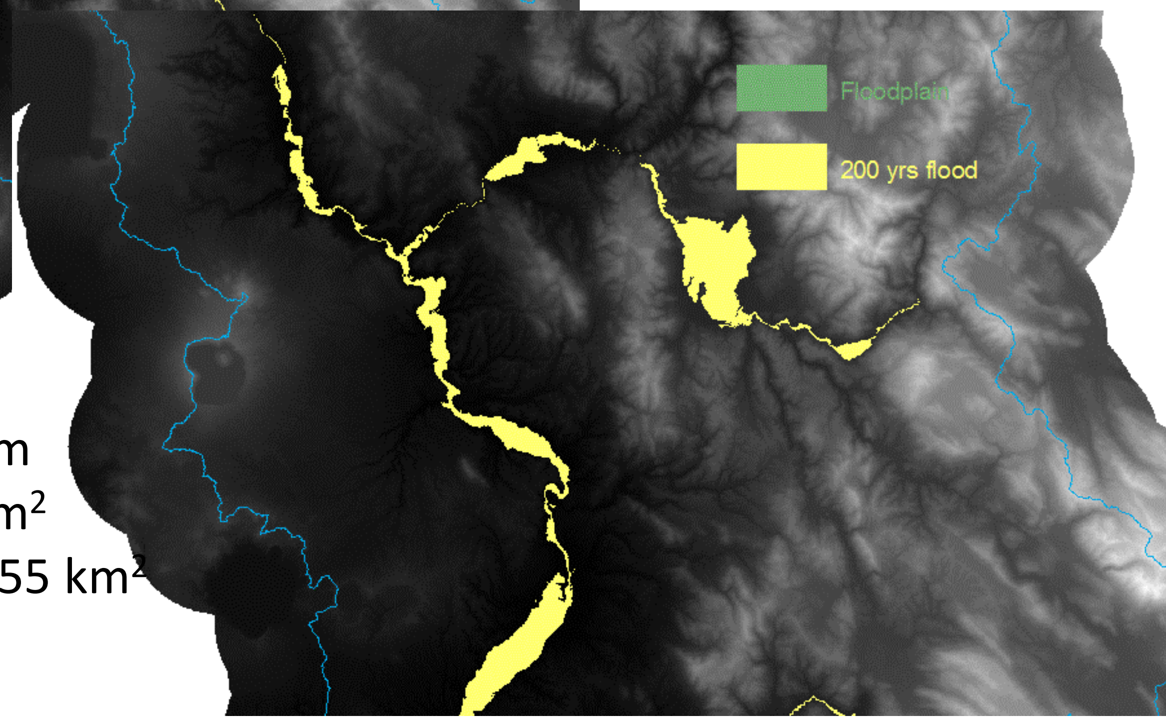
Reference domain	Flood-prone area (km^2) PAI	Floodplain
a	6.08	9.85
b	14.16	18.1
c	13.4	11.1
d	3.21	6.06
e	17.01	13.39
f	15.96	16.22
g	19.41	28.44
h	18.78	15.73

Nardi F., Biscarini C., Di Francesco S., Manciola P. Ubertini L., , Comparing a large scale DEM-based floodplain delineation algorithm with standard flood maps: the Tiber river basin case study, *Journal of Irrigation and Drainage*, 62, (2013)

Tiber river basin

(Central Italy)

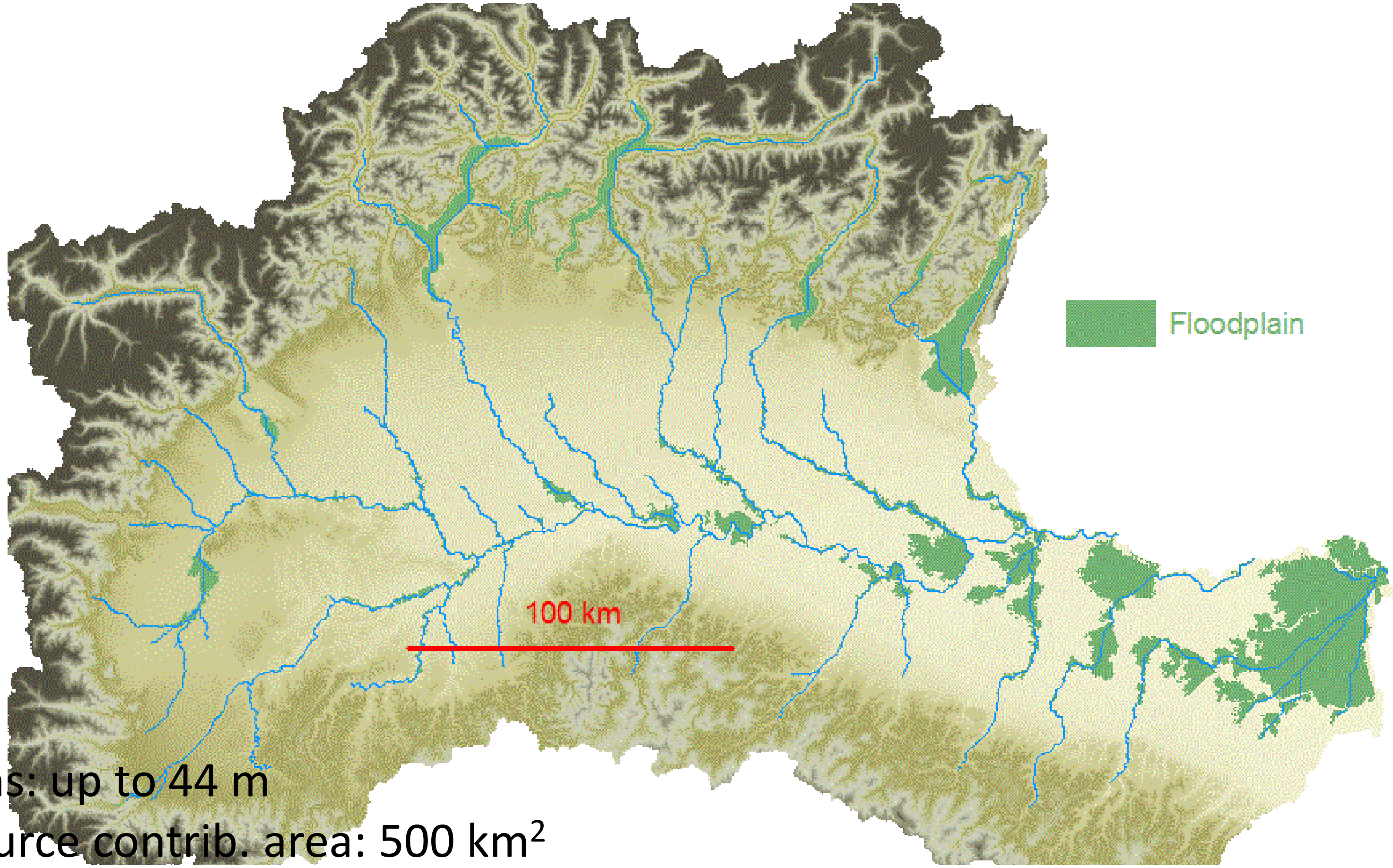




Area 17500 km^2
Channel flow depths: up to 25 m
Stream network source : 500 km^2
Floodplain area: 760 km^2 to 1455 km^2
200 yrs flood maps: 436 km^2

<u>Imposing</u> Leopold law scaling parameters and **exploring** the floodplain method behavior. 64 Combinations a = [0.00005 to 0.002] b = [0.3 to 0.4]

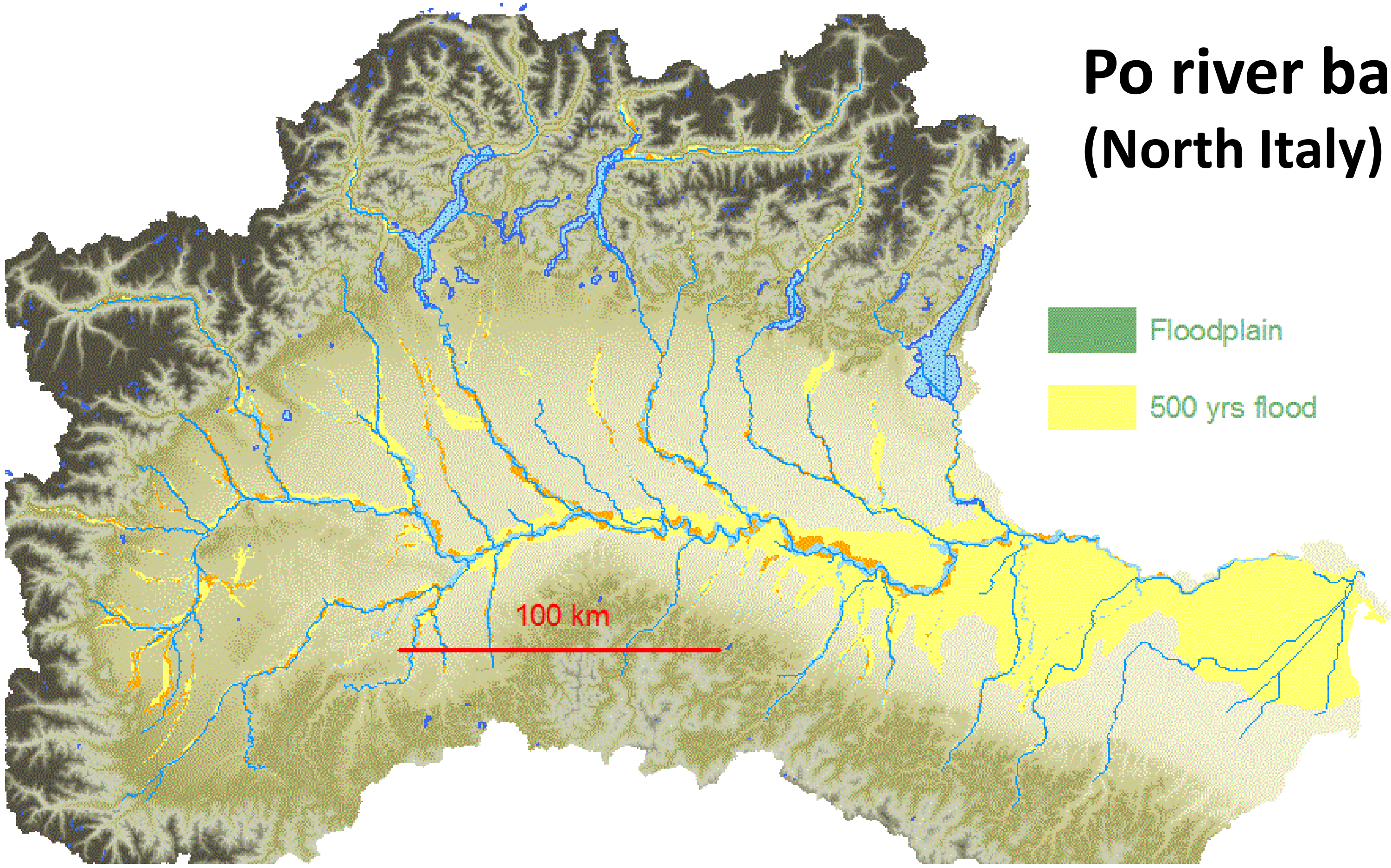
Po river basin (North Italy)



Area 75000 km^2
Channel flow depths: up to 44 m
Stream network source contrib. area: 500 km^2
Floodplain area: 5178 km^2 to 14976 km^2

500 yrs Flood/Inundation maps -> 10,300 km^2

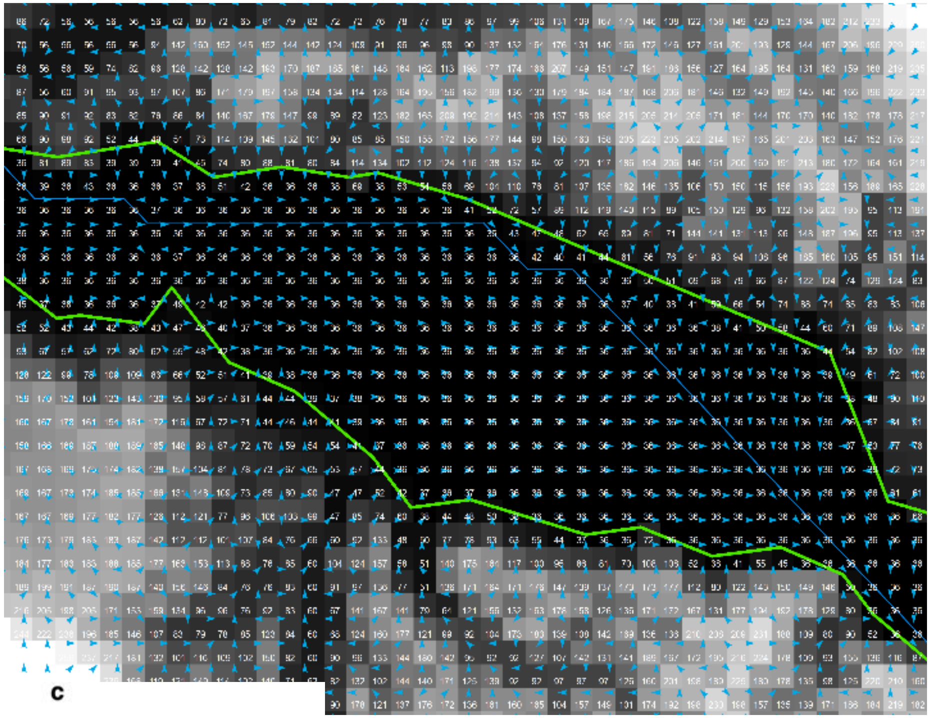
Lake areas: 1000 km^2 Floodplain area: 5178 km^2 to 14976 km^2



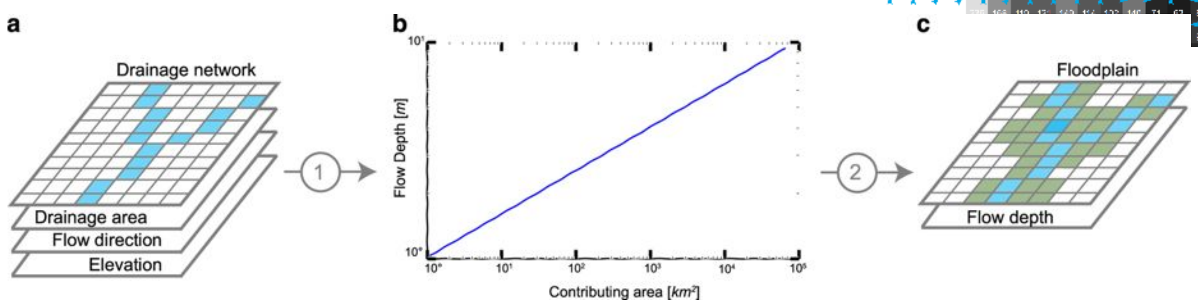
Po river basin
(North Italy)

Global geomorphic floodplain model

- Floodplain extent information, linked to the historical floods (return time 200+ years), is implicitly contained into Digital Terrain Models (DTMs)
- Geomorphic floodplain modelling is based on DTM and on a fluvial flow maximum level h /energy E power law (**h/E ∝ A**) with contributing area A as scaling factor

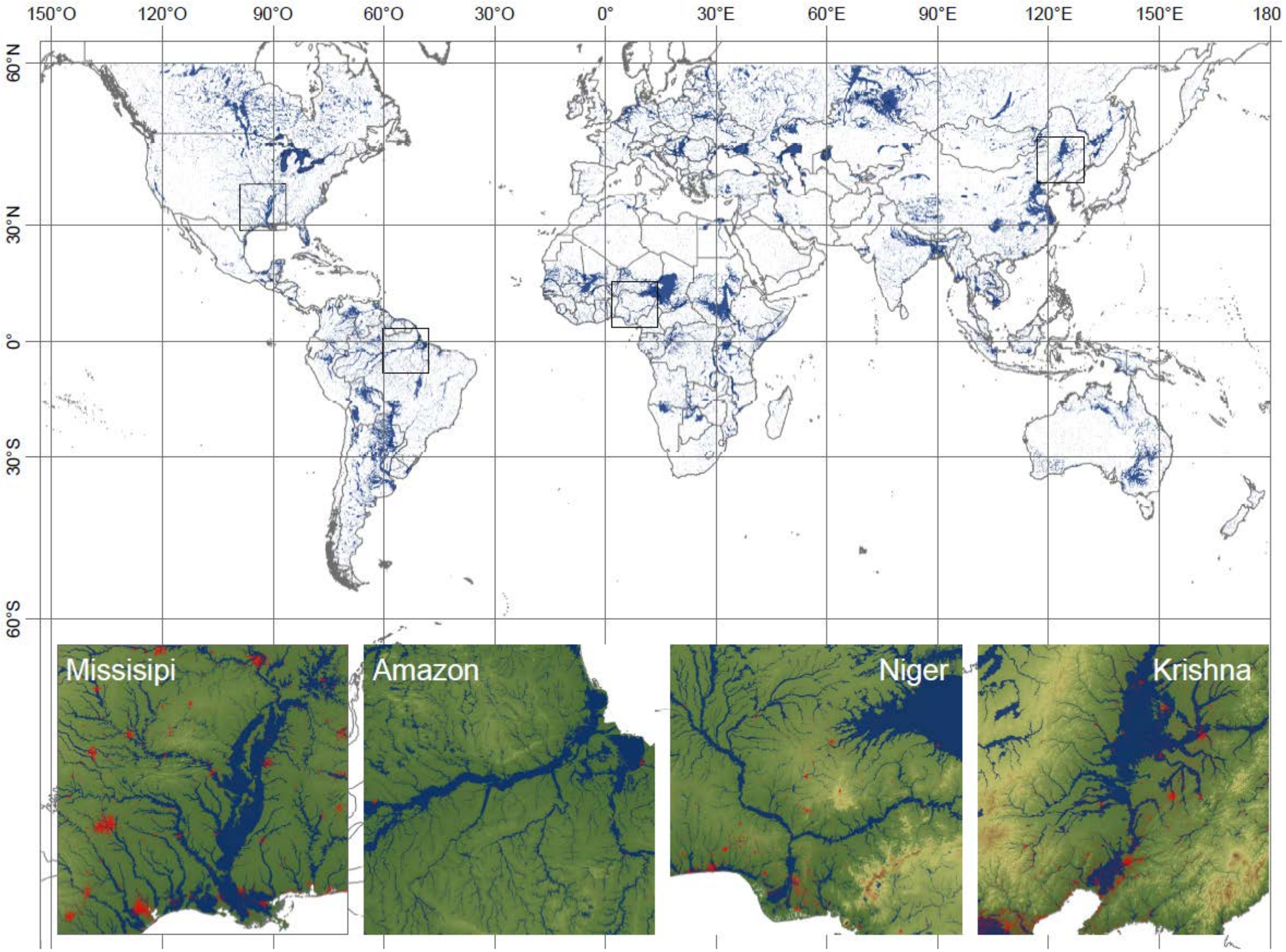


Adapted from
Nardi et al. (WRR, 2006)



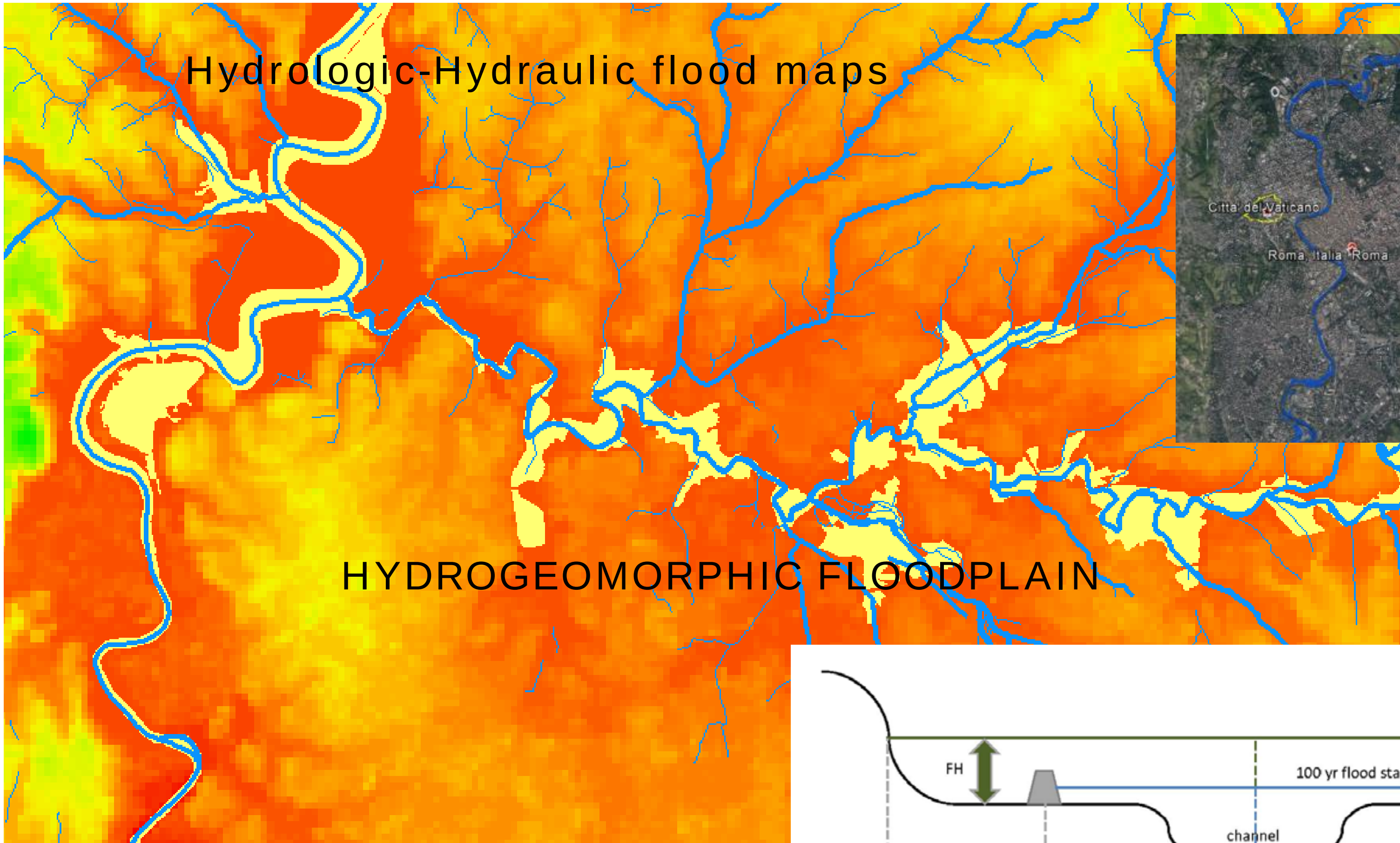
Nardi et al. (SDATA, 2019)

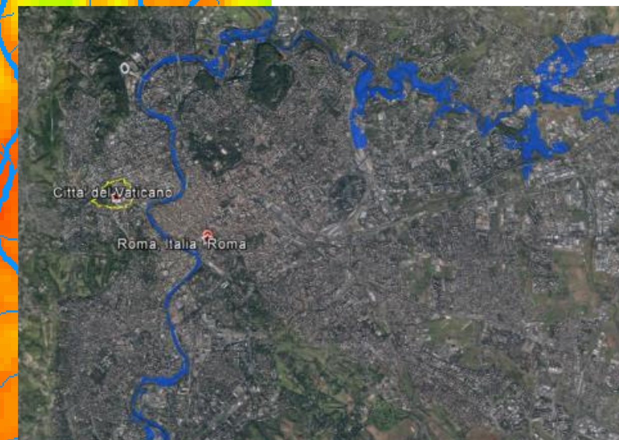
GFPLAIN250m

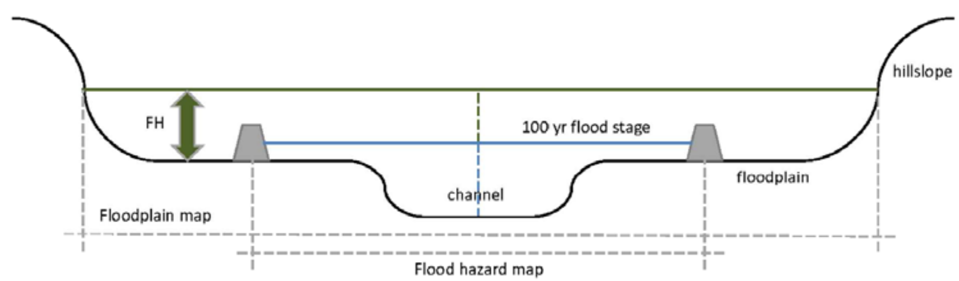


Nardi, F., Annis, A., Di Baldassarre, G., Vivoni, E. R., & Grimaldi, S. (2019). GFPLAIN250m, a global high-resolution dataset of Earth's floodplains. *Scientific data*, *6*, 180309.

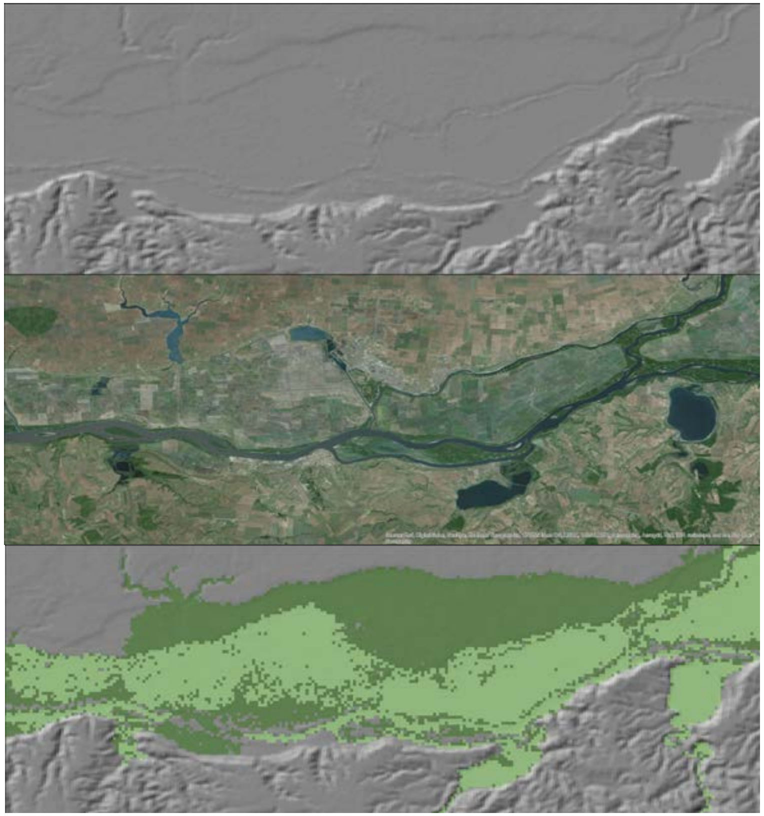
Case studies & regional applications





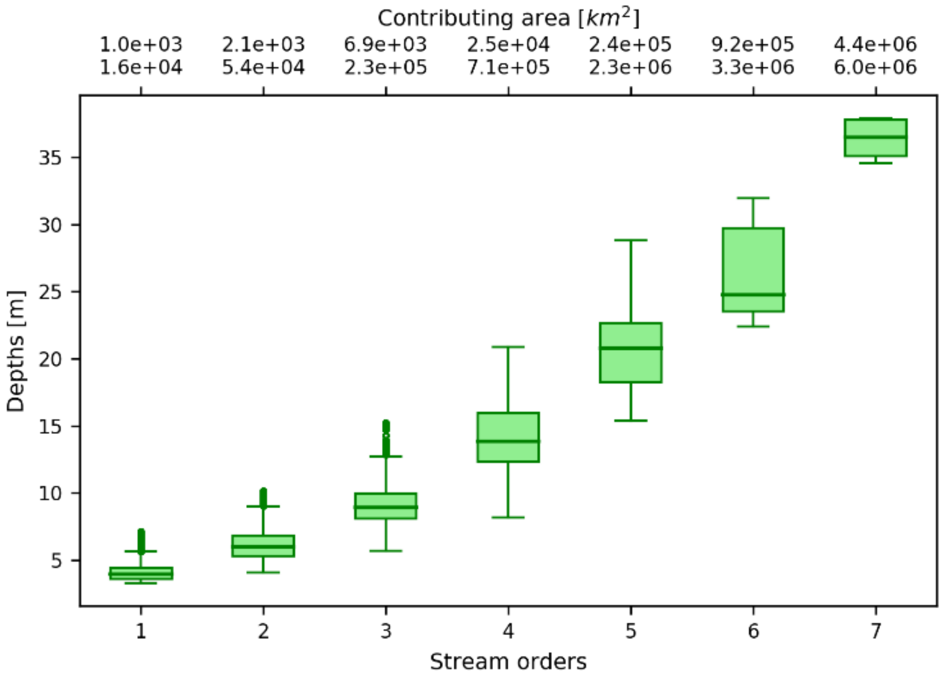


Case studies & regional applications

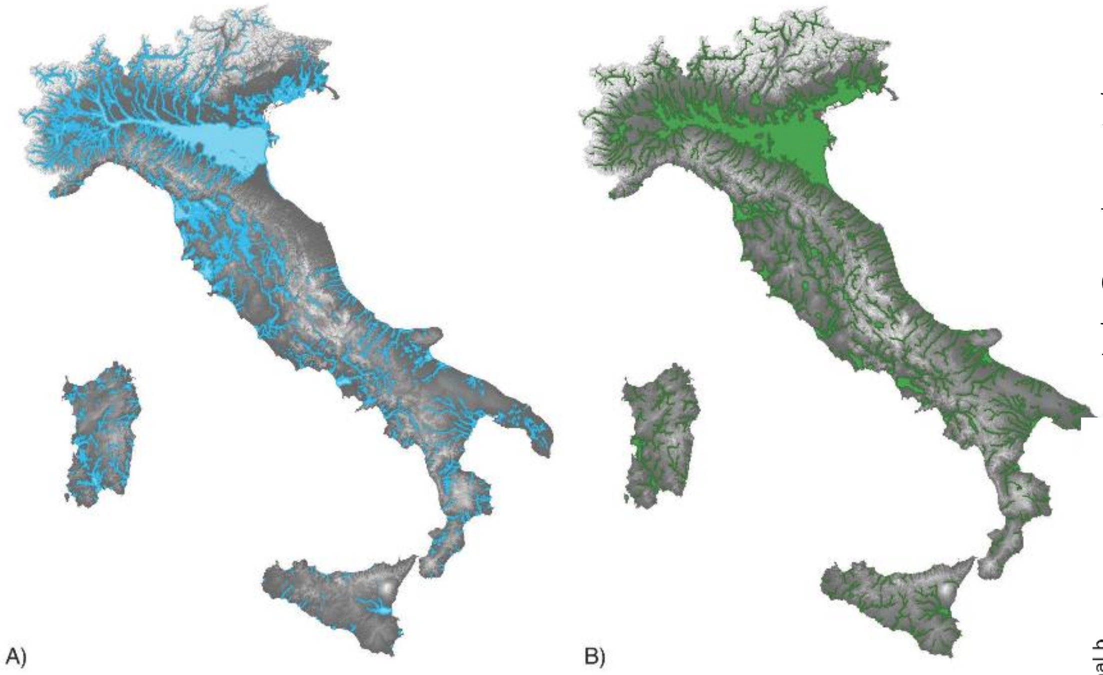


DANUBE

Scaling law parametrization: sensitivity analysis

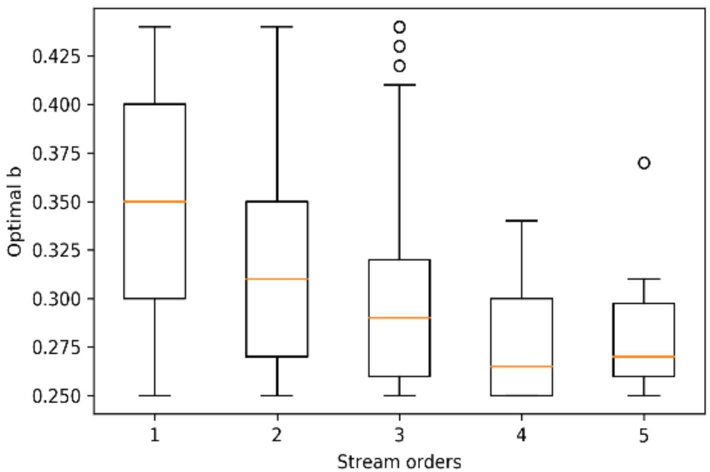


Case studies & regional applications

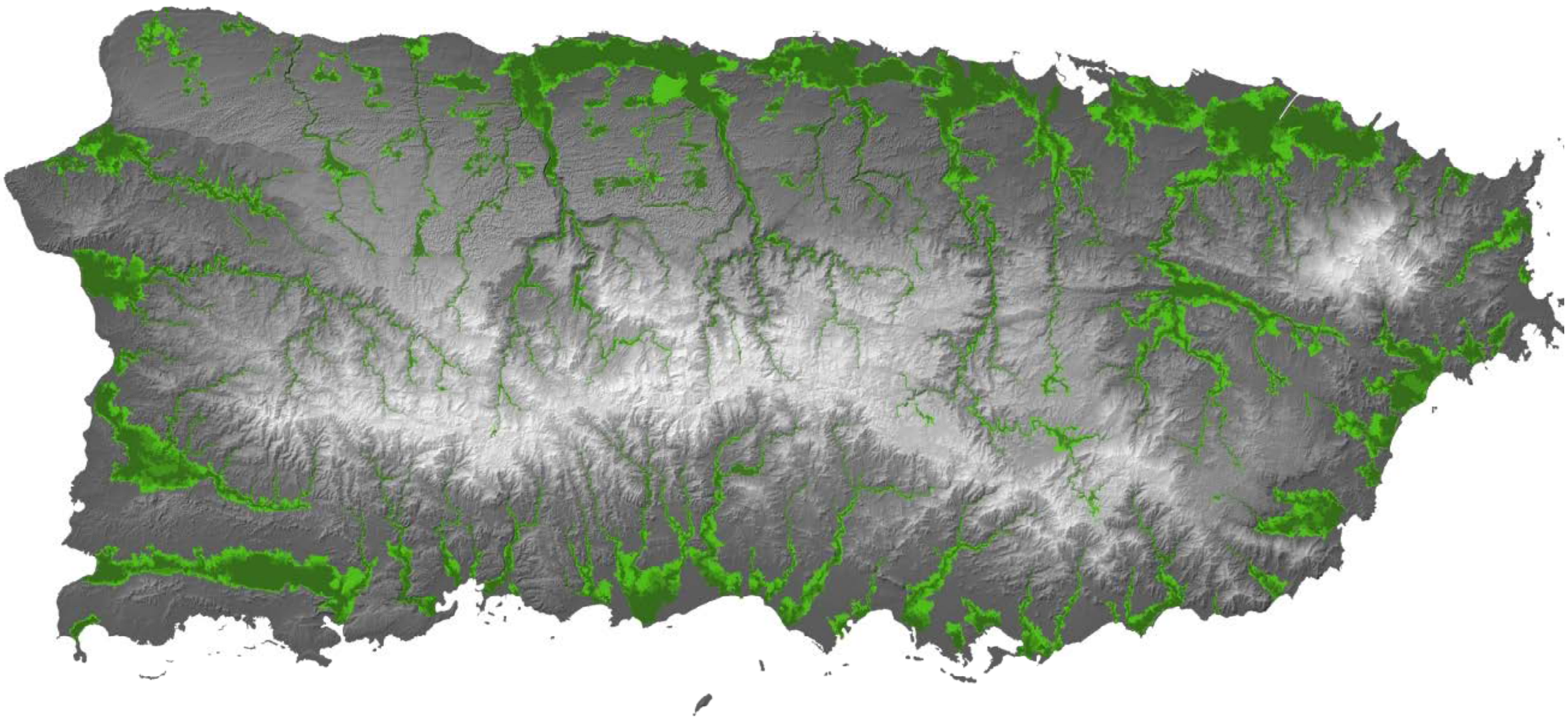


ITALY

Floodplain and flood hazard maps
Consistency/Performance analysis using Measure-to-Fit F indices

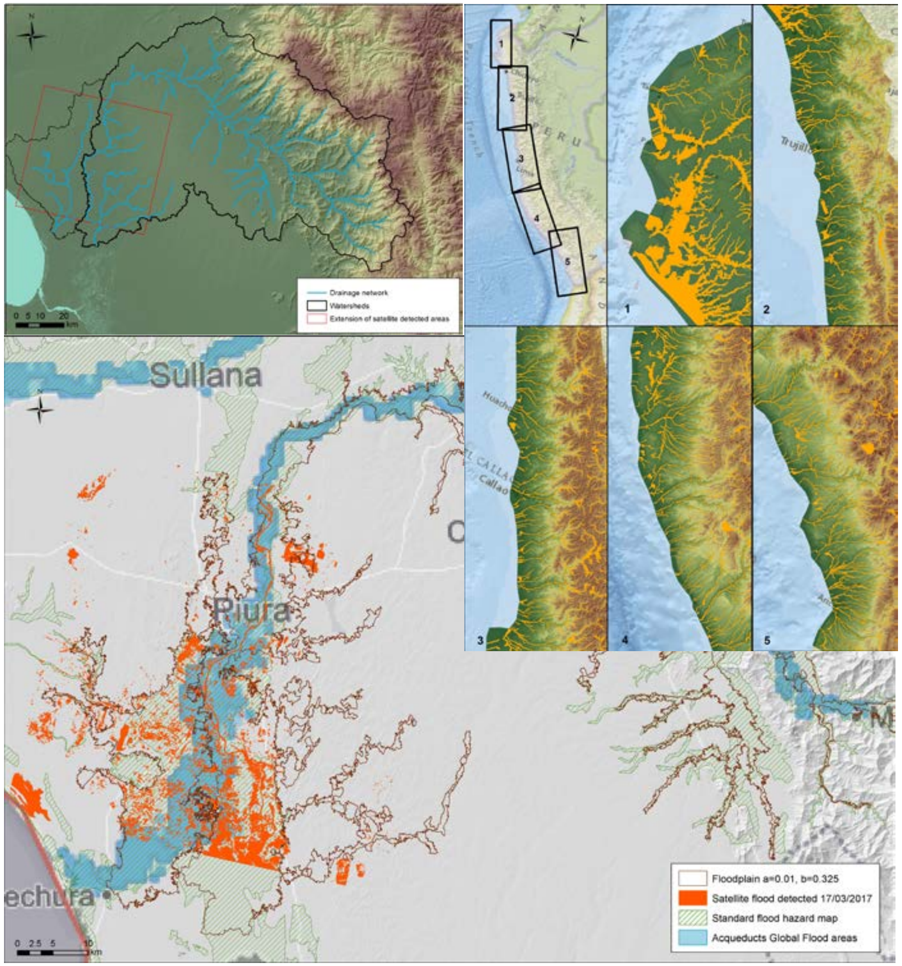


Case studies & regional applications



Coastal floodplain modelling
Puerto Rico

Case studies & regional applications

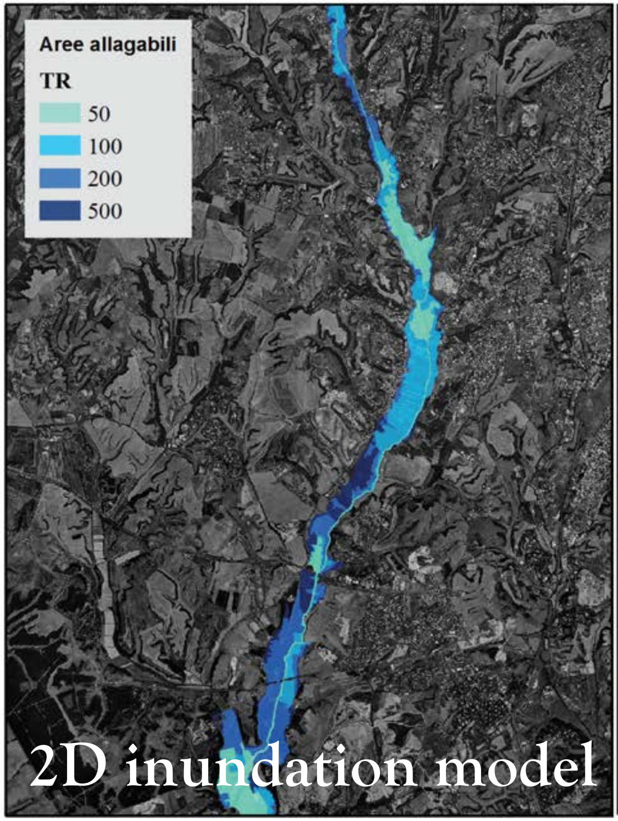


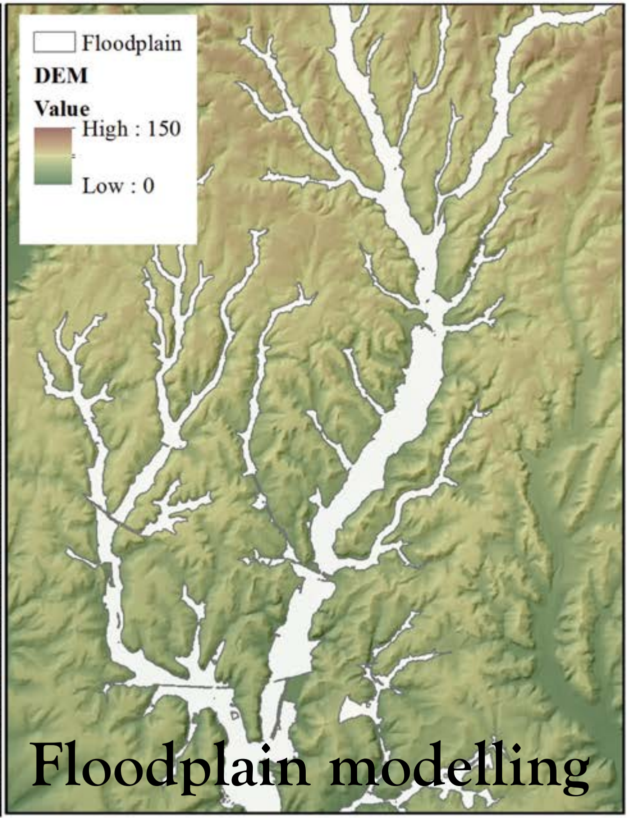
2017 floods in Peru

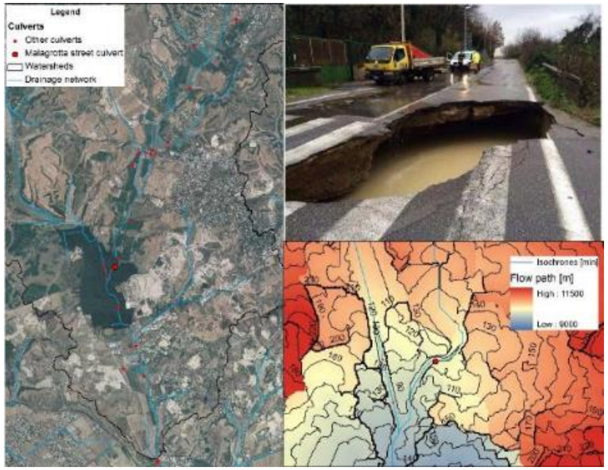
Flood(plain) rapid mapping

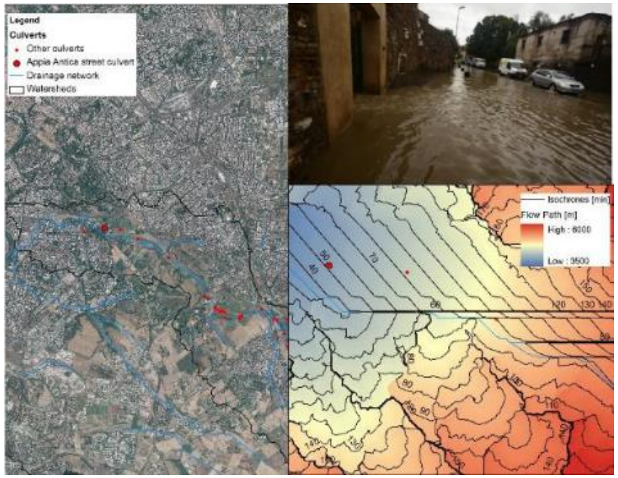
- Satellite-detected flood waters around Piura city, Catacaos and La Union towns in Piura Department, obtained from the Radarsat-2 image acquired on 24 March 2017 (Source UNITAR – UNOSAT)
- Aqueduct flood zoning map

Case studies & regional applications



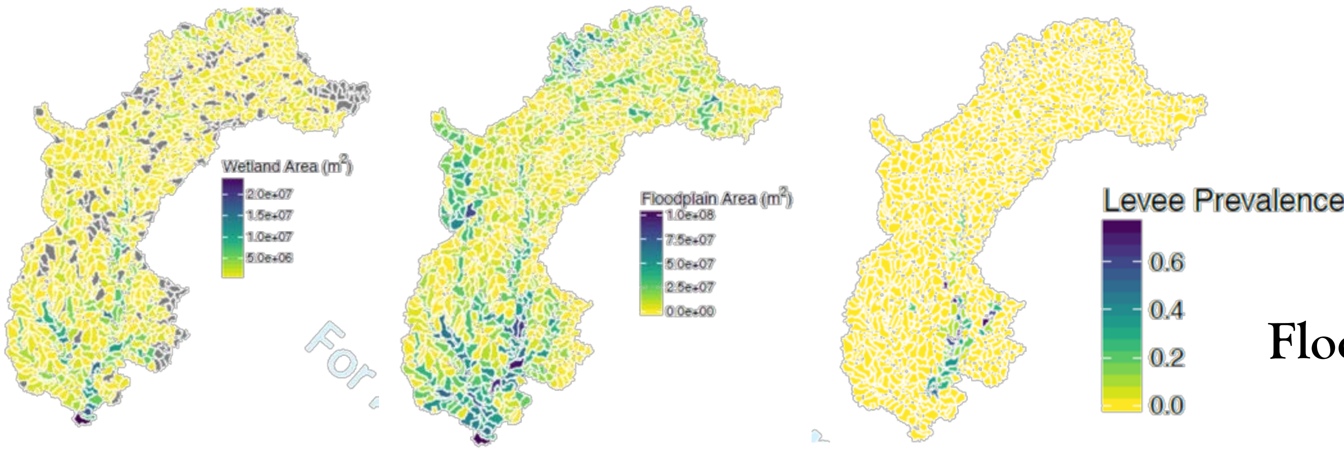




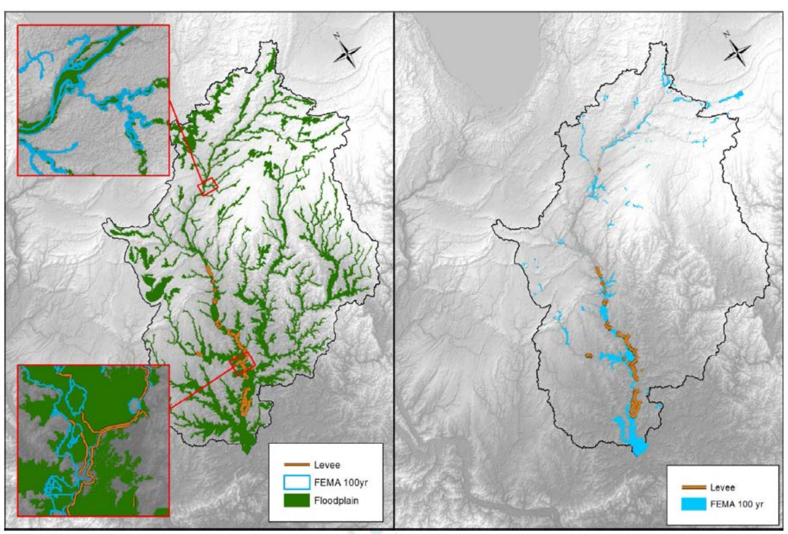


Nardi et al. (Journal of Flood Risk Management, 2018)

Case studies & regional applications



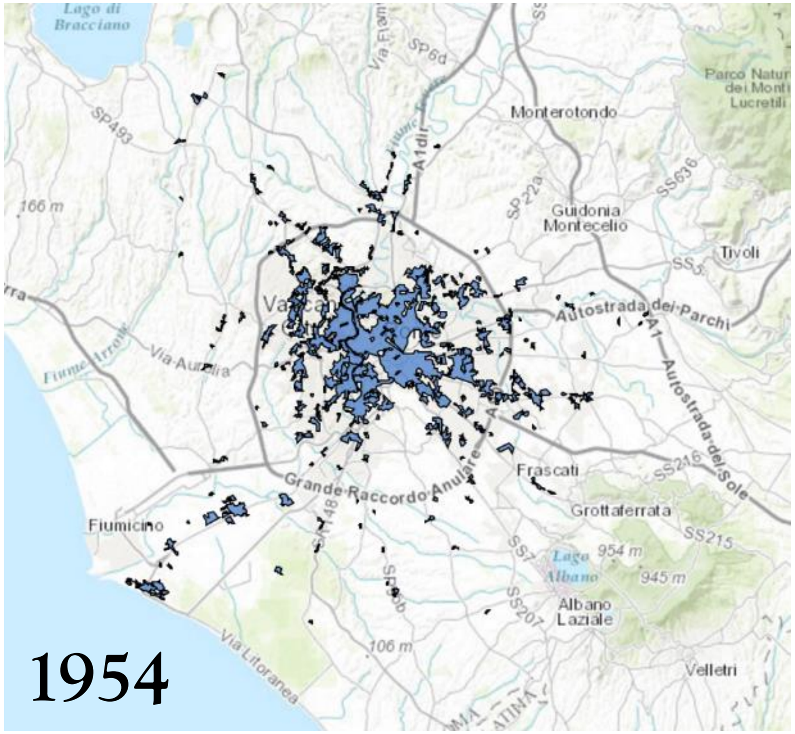
Wabash (USA)
Floodplain disconnectivity

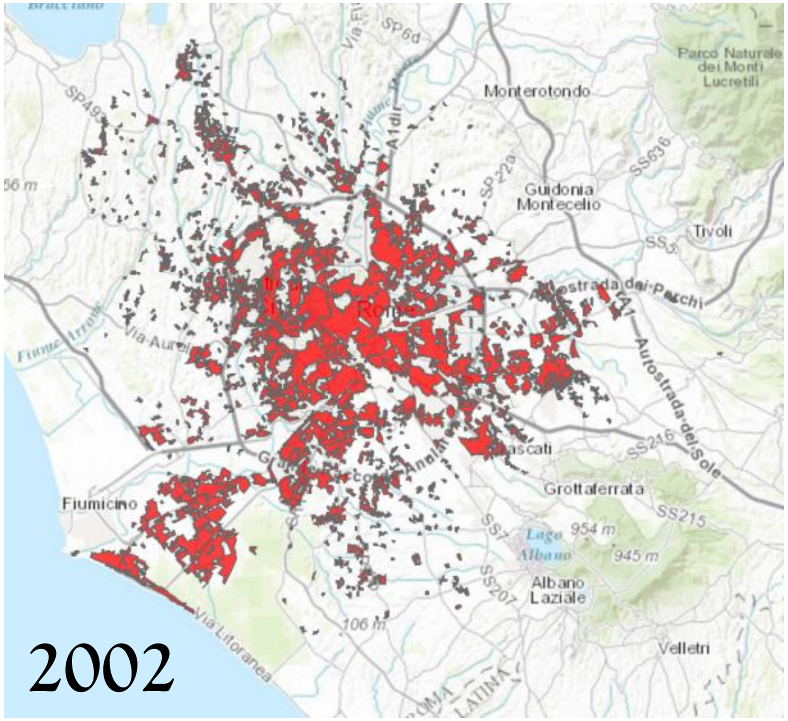


- Tesing geomorphic floodplain model in urbanized basins
- Impact of levees on floodplain/wetland functions

Morrison et al. (JAWRA, 2018)
Nardi et al. (RRA, 2018)

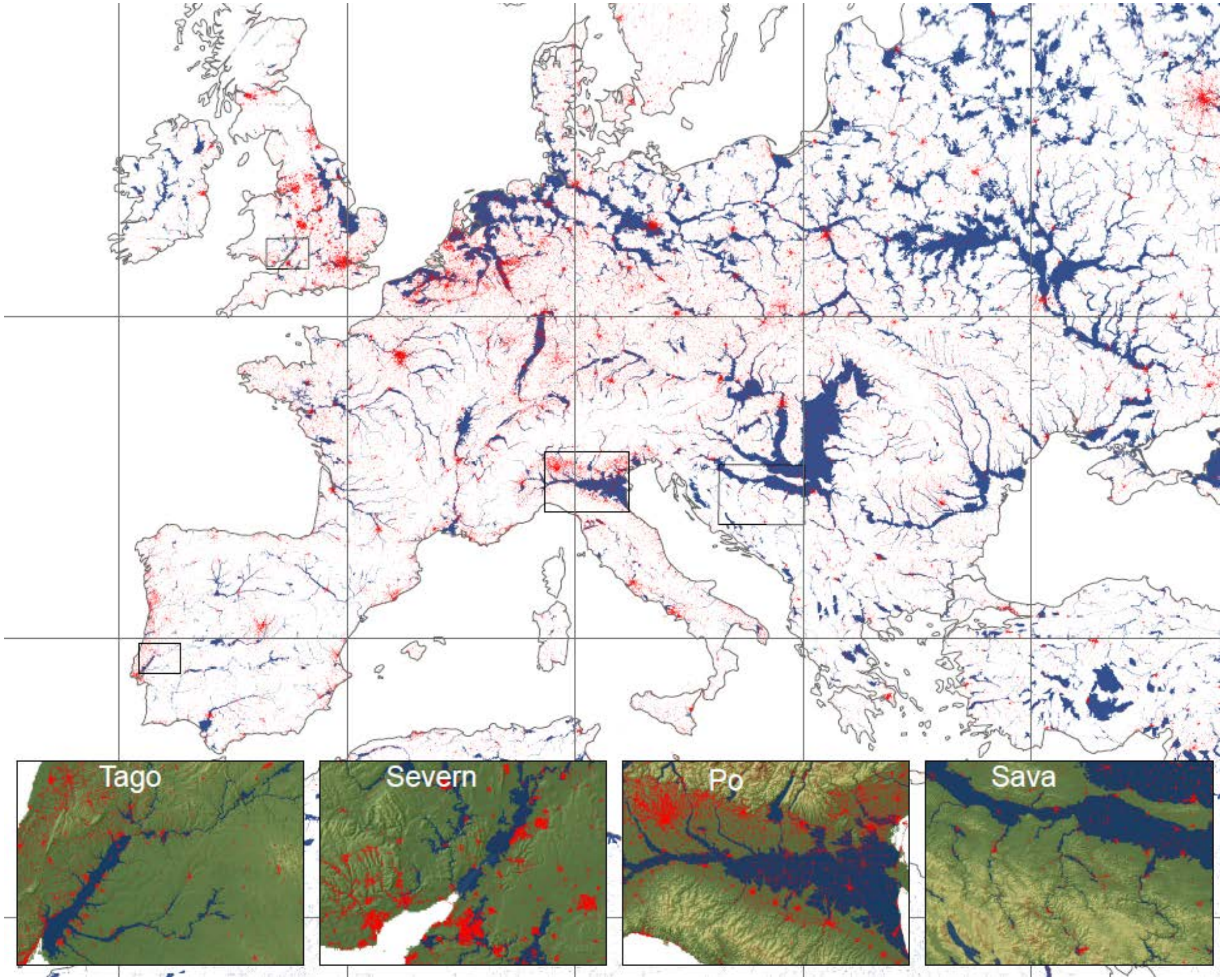
Floods and urban growth





City of Rome

Floods and urban growth



GUF (Global Urban Footprint) data in red color

Floodplains and human impact/behavior/... *

- How many people live in the floodplains? How urban growth has evolved within the floodplain in the last 10, 50, 300 years?
- How much floodplain land (resources? Functions?) was transformed/developed in the last 10-100 years and how much is still available for further human development?
- Floodplain and civilization "next generation model" will be the same as today? Or resource limitations, water risks and other challenges (water quality, pollution, ...) will support a shift towards novel paradigm of urban development?
- **How to develop human dynamics and behavior analysis for floodplain management?**

* *Di Baldassarre, G. D., Kooy, M., Kemerink, J. S., & Brandimarte, L. (2013). Towards understanding the dynamic behaviour of floodplains as human-water systems. Hydrology and Earth System Sciences, 17(8), 3235-3244.*

References

- Ward, P. J., Jongman, B., Salamon, P., Simpson, A., Bates, P., De Groeve, T., ... & Winsemius, H. C. (2015). Usefulness and limitations of global flood risk models. Nature Climate Change, 5(8), 712.
- Nardi, F., Vivoni, E. R., & Grimaldi, S. (2006). Investigating a floodplain scaling relation using a hydrogeomorphic delineation method. Water Resources Research, 42(9).
- Nardi, F., Biscarini, C., Di Francesco, S., Manciola, P., & Ubertini, L. (2013). Comparing a large-scale DEM-based floodplain delineation algorithm with standard flood maps: The Tiber River Basin case study. Irrigation and Drainage, 62(S2), 11-19.
- Nardi, F., Annis, A., & Biscarini, C. (2018). On the impact of urbanization on flood hydrology of small ungauged basins: The case study of the Tiber river tributary network within the city of Rome. Journal of Flood Risk Management, 11, S594-S603.
- Morrison, R. R., Bray, E., Nardi, F., Annis, A., & Dong, Q. (2018) Spatial Relationships of Levees and Wetland Systems within Floodplains of the Wabash Basin, USA. Journal of the American Water Resources Association (JAWRA) 54(4): 934–948
- Nardi, F., Morrison, R. R., Annis, A., & Grantham, T. E. (2018). Hydrologic scaling for hydrogeomorphic floodplain mapping: Insights into human-induced floodplain disconnectivity. River Research and Applications, 34(7), 675-685.

Thanks for the attention

Università per Stranieri di Perugia

http://warredoc.unistrapg.it

fernando.nardi@unistrapg.it

Citizen Science and Big Data in Hydrology

Part 2. Introducing citizen science and big data: a general overview and some studies

January 31, 2019

Fernando Nardi

Which is the definition of "ungauged basin"?

Is this geospatial dataset / flow observation / time series / .. "accurate"? Are those data good enough?

... ask yourself these questions now, and in 5-10 years?

The concept of data quality/quantity is being completely re-defined in the era of data rich hydrology..
... but information is what we need

Can we (have to) put humans into the equations?

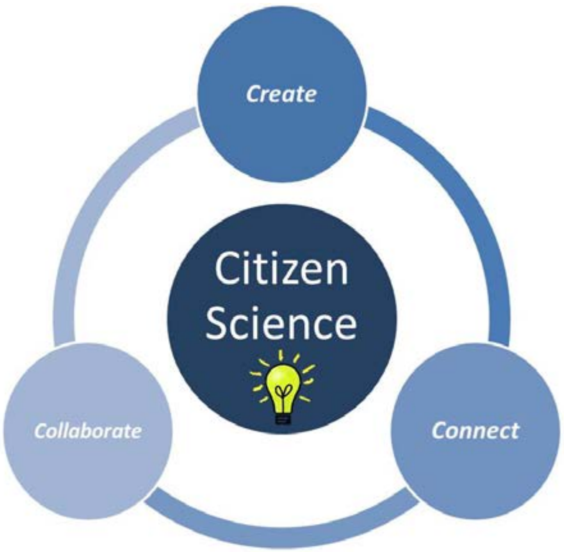
Let's start with some **DEFINITIONS**

Citizen Science
Active Citizenship
Open Data
Big Data
…

Citizen Science

Citizen science describes the engagement of people in scientific processes who are not tied to institutions in that field of science. Participation can range from the short-term collection of data to the intensive use of leisure time in order to delve deeper into a research topic together with scientists and/or other volunteers. Although many volunteer scientists do have a university degree, this is not a prerequisite for participating in research projects. However, it is important that scientific standards are adhered to.

Source: [Zentrum fur Citizen Science]

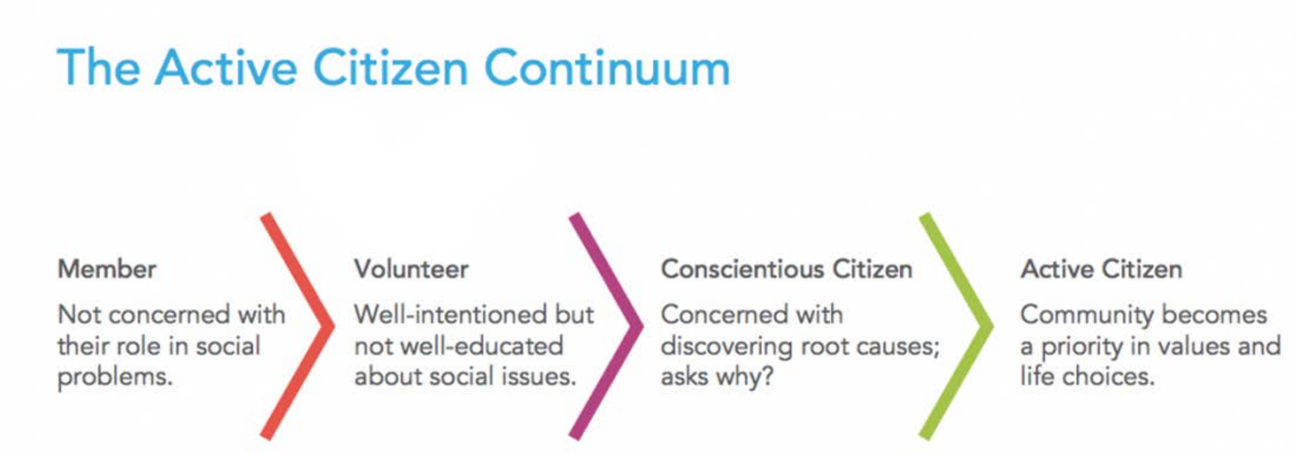


Source: [South Atlantic Fishery Management Council]

Active Citizenship

Active citizenship means people getting involved in their local communities and democracy at all levels, from towns to cities to nationwide activity. Active citizenship can be as small as a campaign to clean up your street or as big as educating young people about democratic values, skills and participation. Active citizenship is one of the most important steps towards healthy societies especially in new democracies

Source: *Open Society Foundations*



Source: *Center for Service-Learning and Civic Engagement – University of Kentucky*

Crowdmapping

The term Crowdmapping is composed of the two words Crowd = Large number of people gathered together and Mapping = Map / Cartography and was introduced by the theme of emergency management (specifically during natural disasters) to publish information by the high amount of users in an area producing messages, information, images and video. This information allow the creation of Digital Map (primarily online and in real time) thanks to the application of smart phones and web platforms developed specifically for disaster management.

Source: [Crowdmap]

Open Data

Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.

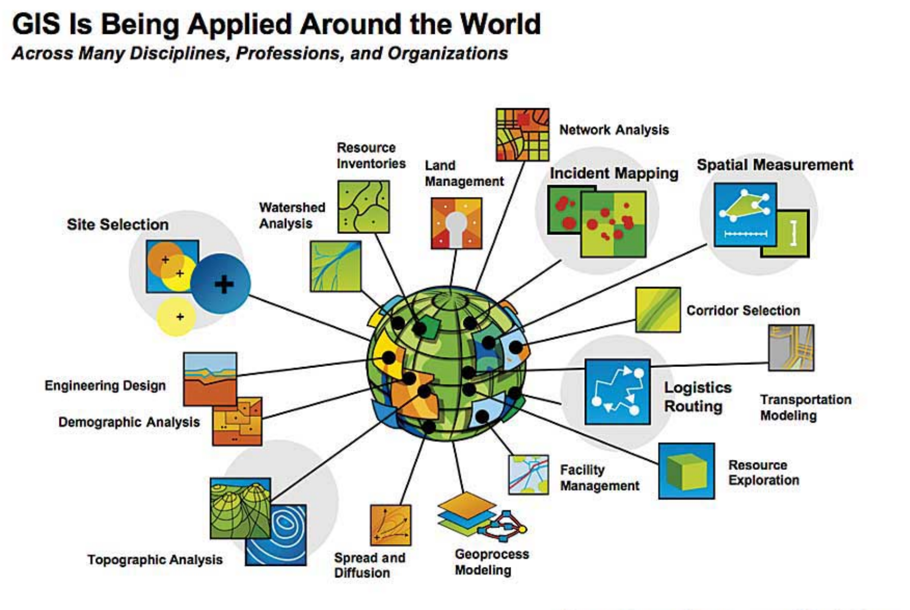
Availability and Access: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
Re-use and Redistribution: the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.
Universal Participation: everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

Source: *Open Knowledge International*

Source: *Open Knowledge International*

Geographic Information System (GIS)

GIS (Geographic Information System) is a ICT-based system for management, analysis and visualization of geospatial data. GIS consists of users, software / hardware, procedures, methods and tools that characterized the built environment (or platform) for gathering and producing geographic information.

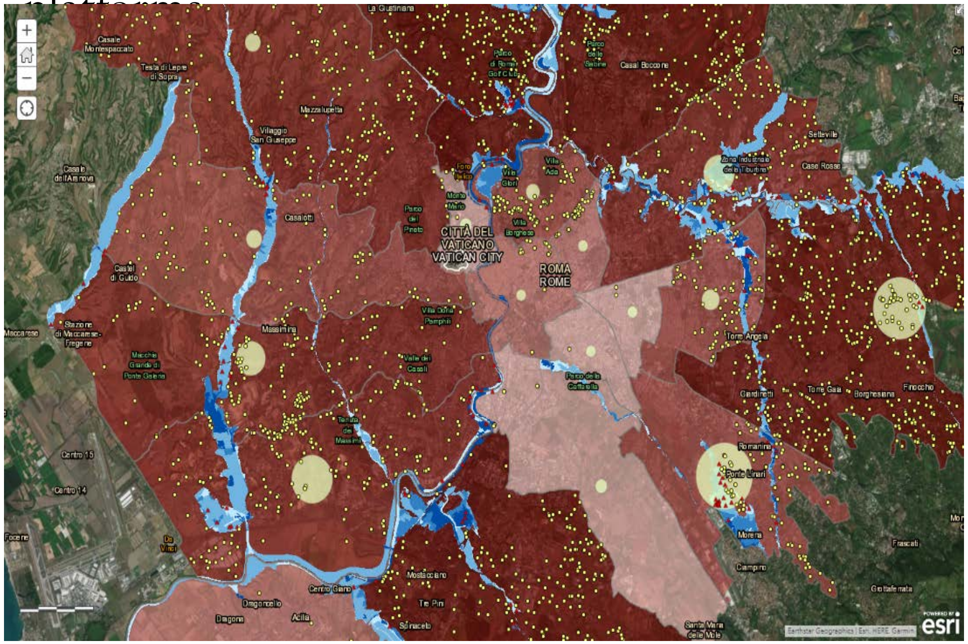


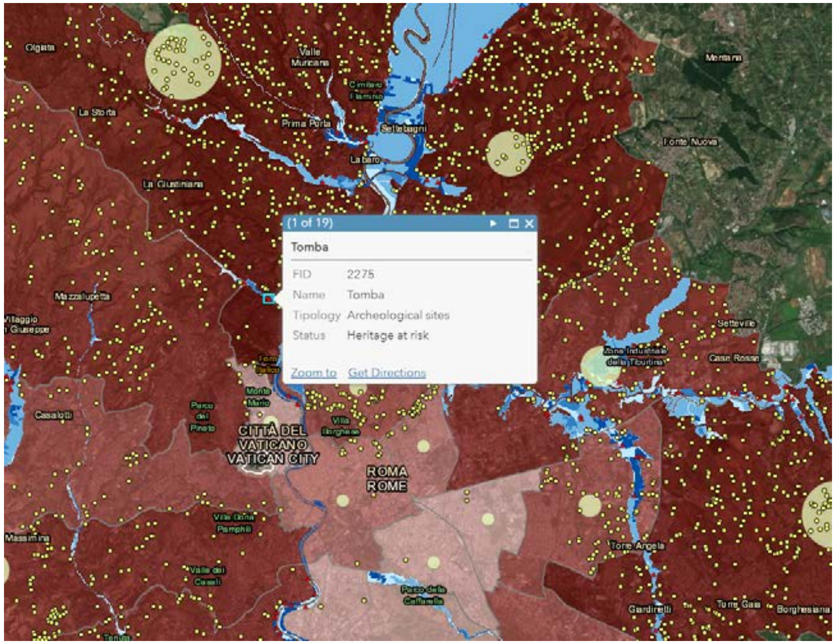
Source: ESRI

A geographic information system (GIS) can be defined as a computer application capable of performing virtually any conceivable operation on geographic information, from acquisition and compilation through visualization, query, and analysis to modeling, sharing, and archiving (Goodchild, 1993)

Geospatial maps (GIS) & Digital mapping

Digital Mapping refers to the use and production of geographic information to create digital maps and virtual images in the form of maps whose content can be inquiry and directed in a dynamic and interactive way by the user. The interaction between the Digital Map and the user can be done through dedicated software that is installed on your computer (desktop application) using the Web browser, commonly used for browsing, such as web based cloud platforms.





Geospatial tools and data are being redefined with new cloud-based platforms

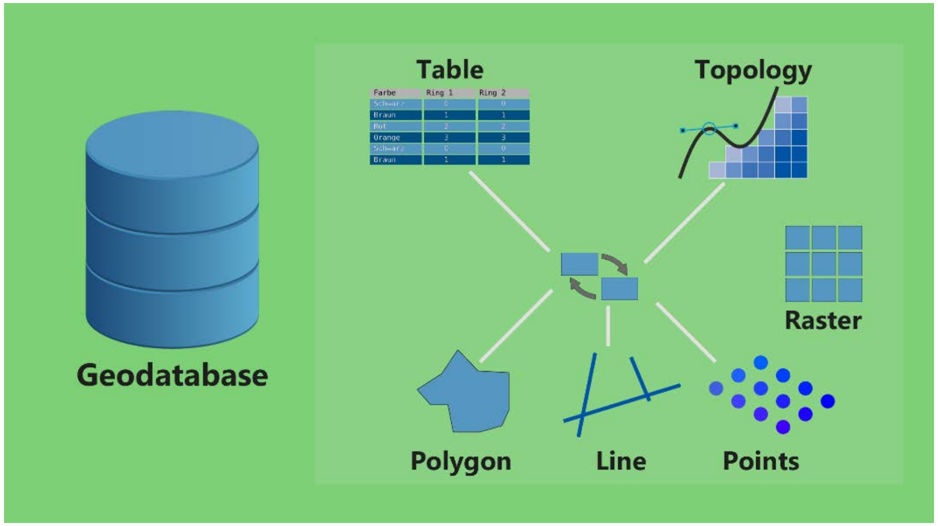
Geodata

Any dataset where data points include a location, e.g. as latitude and longitude or another standard encoding. Maps, transport routes, environmental data, cadastral data, and many other kinds of data can be published as geodata.

Source: Open Knowledge International

Geodata is information about geographic locations that is stored in a format that can be used with a geographic information system (GIS). Geodata can be stored in a database, geodatabase, shapefile, coverage, raster image, or even a dbf table or Microsoft Excel spreadsheet.

Source: ESRI – ArcGIS for Desktop



Source: GIS Geography

Volunteered Geographic Information (VGI)

- **Volunteered geographic information (VGI)** is the harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals. VGI is a special case of the larger Web phenomenon known as user-generated content, and allows citizens to have a more active role in urban planning activities.
- One of the most important elements of VGI in contrast to standard UGC is the geographic element, and its relationship with collaborative mapping. The information volunteered by the individual is linked to a specific geographic region. While this is often taken to relate to elements of traditional cartography, VGI offers the possibility of including subjective, emotional, or other non-cartographic information.
- Geo-referenced data produced within services such as TripAdvisor, Flickr, Twitter, Instagram and Panoramio can be considered as VGI.

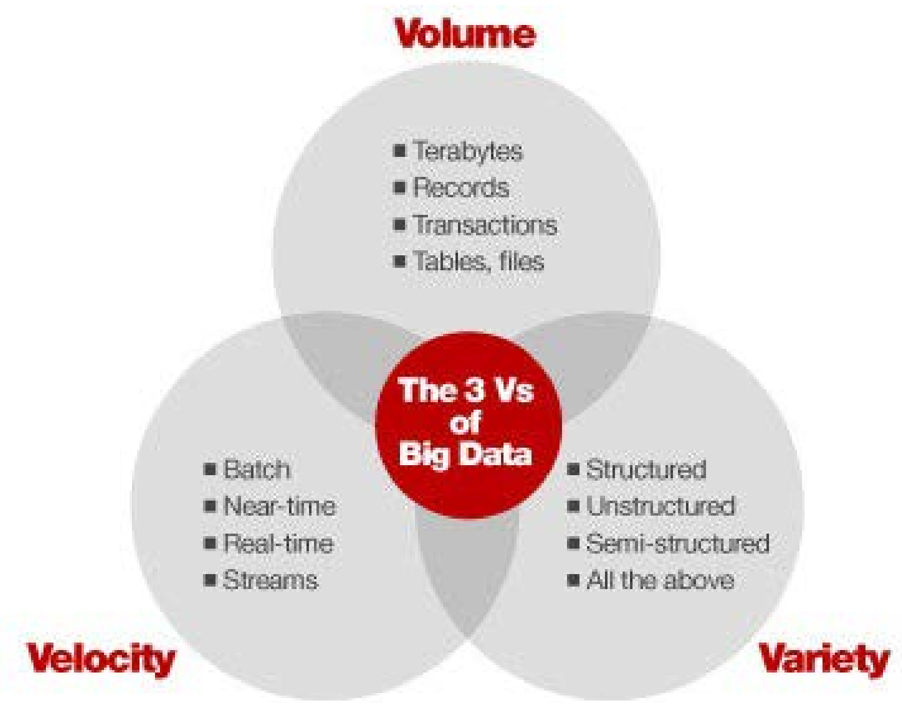
M.F. Goodchild (2007), *Citizens as sensor: the world of volunteered geography*, GeoJournal, 69, pp. 211-221

Source: vgiscience.org

Big data

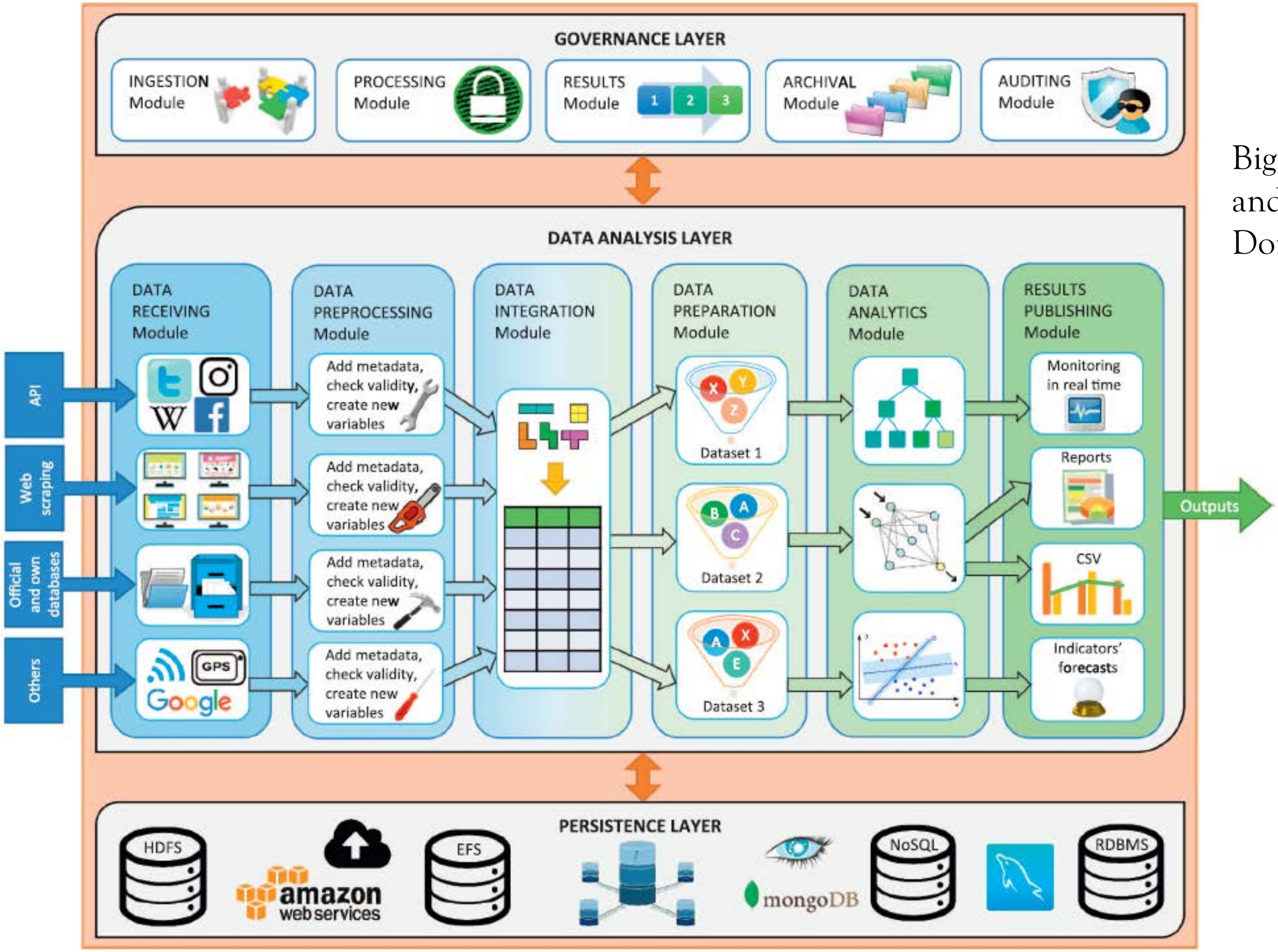
Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. But it's not the amount of data that's important. It's what organizations do with the data that matters. Big data can be analyzed for insights that lead to better decisions and strategic business moves.

Source: [SAS the power to know]



Source: [Analytics Training]

Big Data architecture and work flow



Big Data architecture for nowcasting and forecasting (Blazquez & Domenech, 2018)

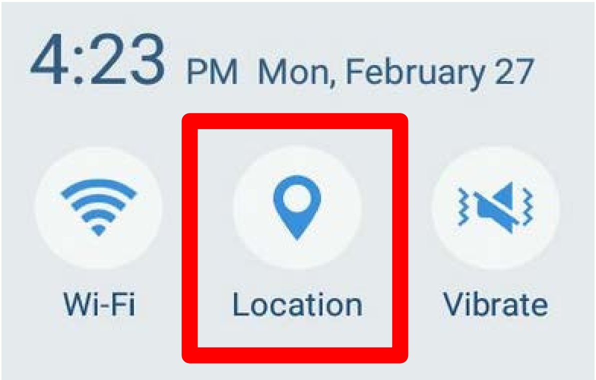
Introduction to Big data

- First time «Big Data» was mentioned at the end of the '90 (Cox & Ellsworth, 1997) referring to the 3V: *Volume* (data volumes), *Velocity* (data transfer/production rate) e *Variety* (diverse type of data).
- Today we use the 5V, adding the Value *(data with Value)*, *Veracity* (meaningful and properly managed data) (Bello-Orgaz et al., 2016)
- Raw data resulting from satellite, aerial images, surveys, polls or collection is processed and structured to be readable and allow the creation of new data.
- Data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day.
- Data can arrive from real-time measuring instruments, such as video, audio, web-based extensible markup language (XML), or any of a plethora of other media formats, and be processed and delivered in near or actual real time, through distributed environments: Mobile devices, Tablets, Sensor-based-internet-devices
- We are becoming individual sensors (Receiving and transmitting information in real time). A huge sensor network of people is created by simple holding a mobile device.

Data collection

- Data can be collected
  - Openly from social media platforms and crowdsourced mechanisms (e.g. OpenStreetMap, Ushahidi, etc.).
  - Covertly (without our consent or knowledge)
- While the information harvested from social media is not explicitly geospatial, it does include implicit geospatial content, rendering it suitable for new types of geospatial analysis. For example, Twitter content that includes geotags may serve as a timely, fairly accurate reporting system for natural disasters, with human beings acting as the sensors that collect and report the information.

Source: image.flaticon.com

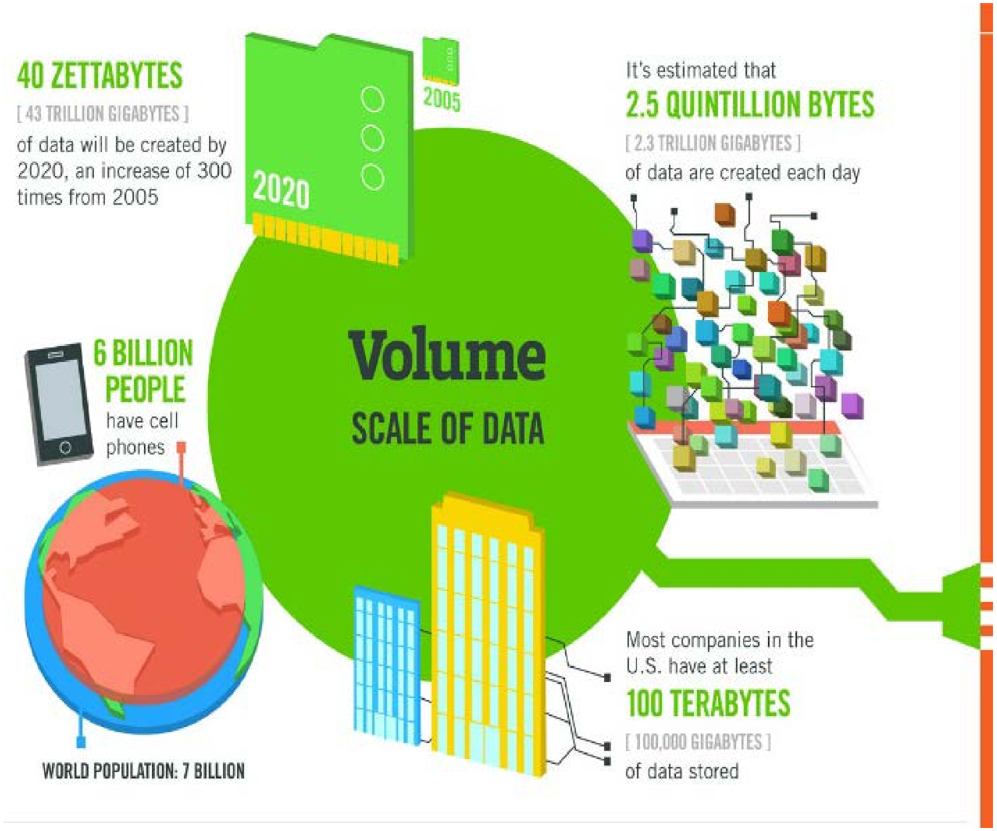


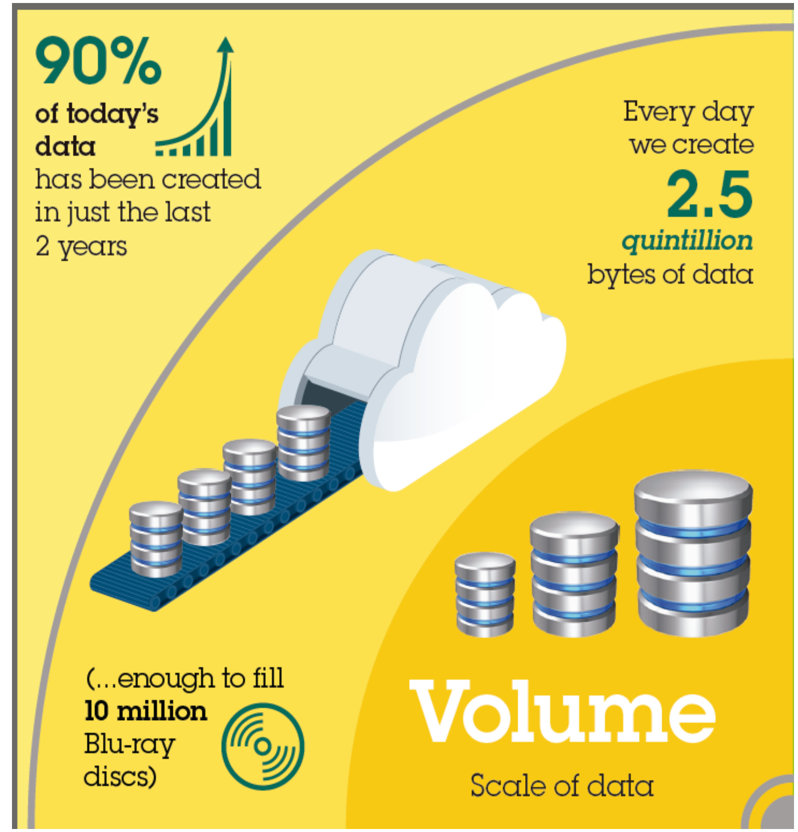
Source: Samsung Galaxy G7

Structured and non-structured Big Data

- Big Data are classified as:
  - *structured* (tabular data generally organized in relational databases (RDBMS)
  - *non structured*, (images, video, social media posts, … everything else that is not structured).
- 95% of Big Data is generally non structured

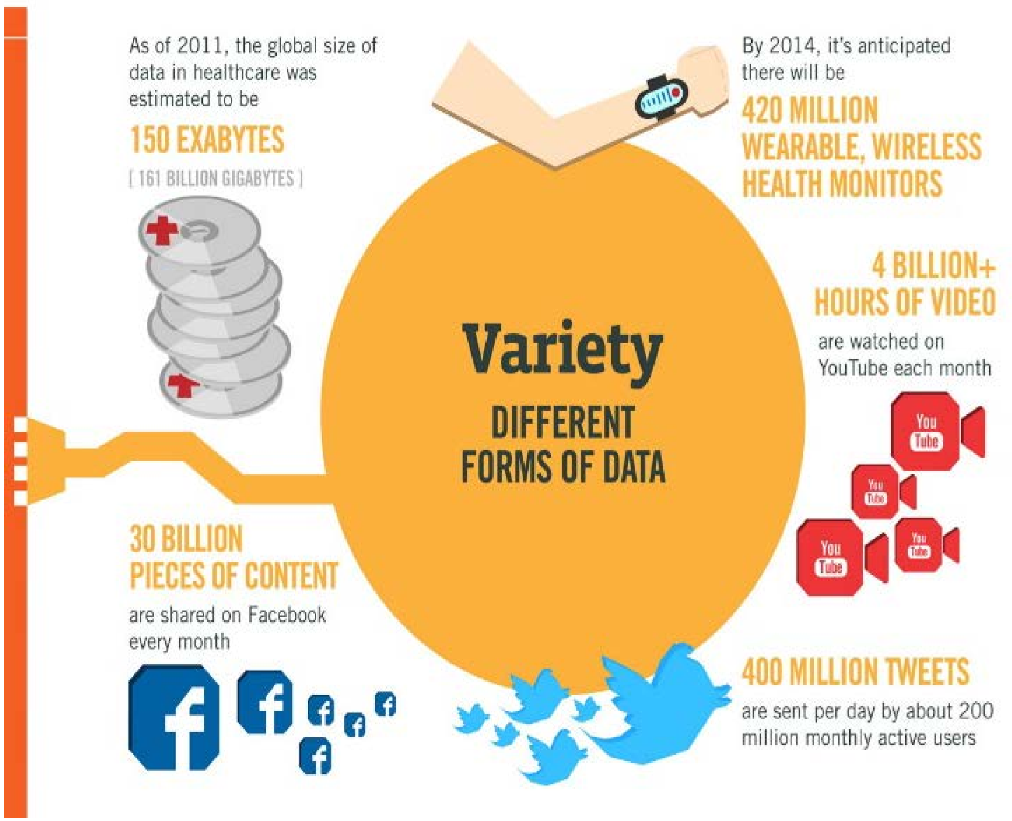
The Four Dimensions of Big Data: Volume

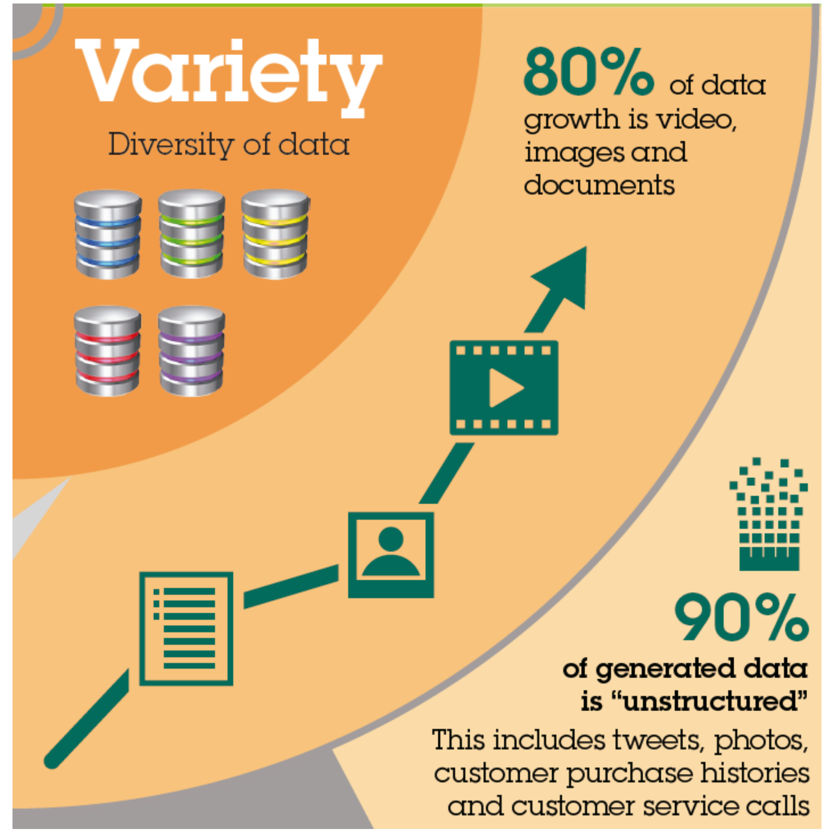




Source: www.ibmbigdatahub.com/infographic

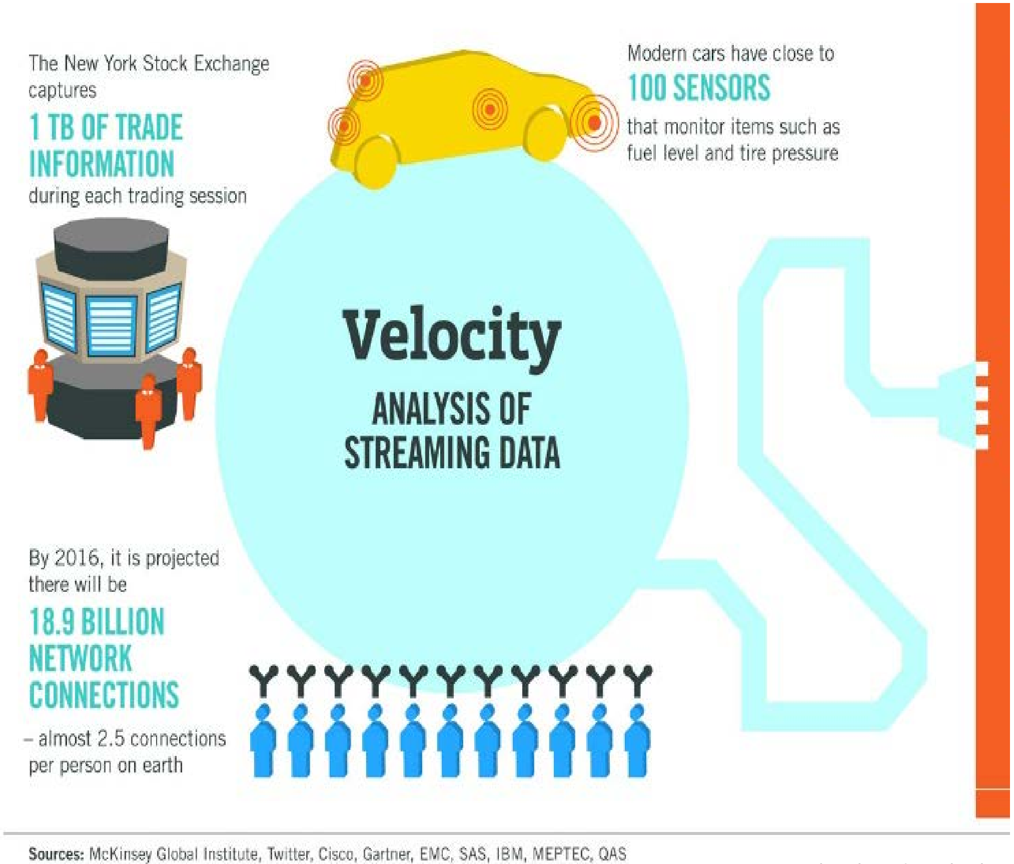
The Four Dimensions of Big Data: Variety

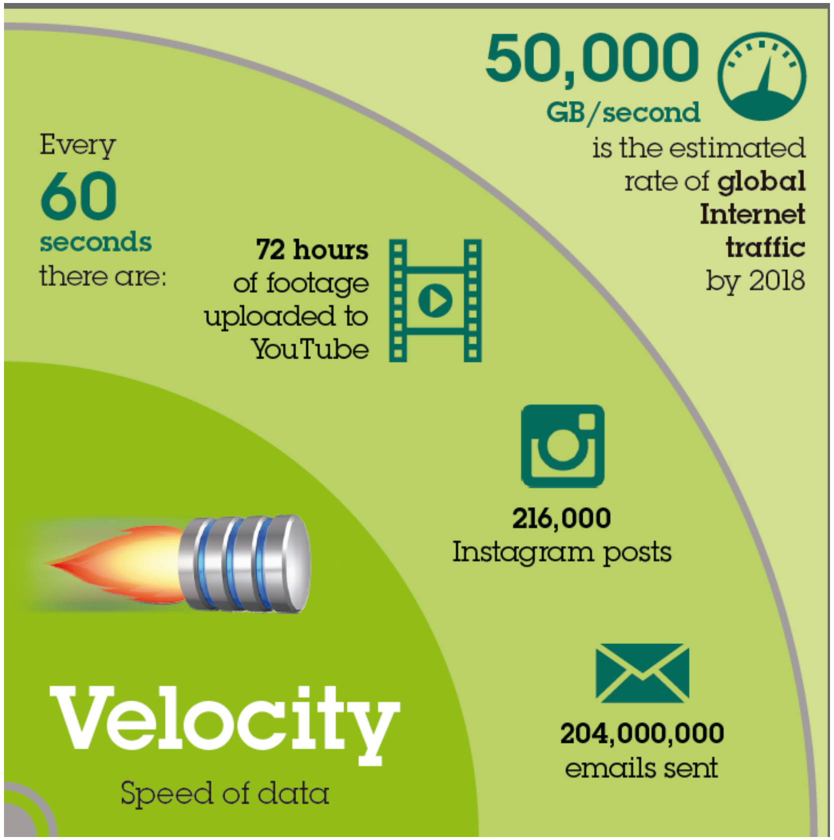




Source: www.ibmbigdatahub.com/infographic

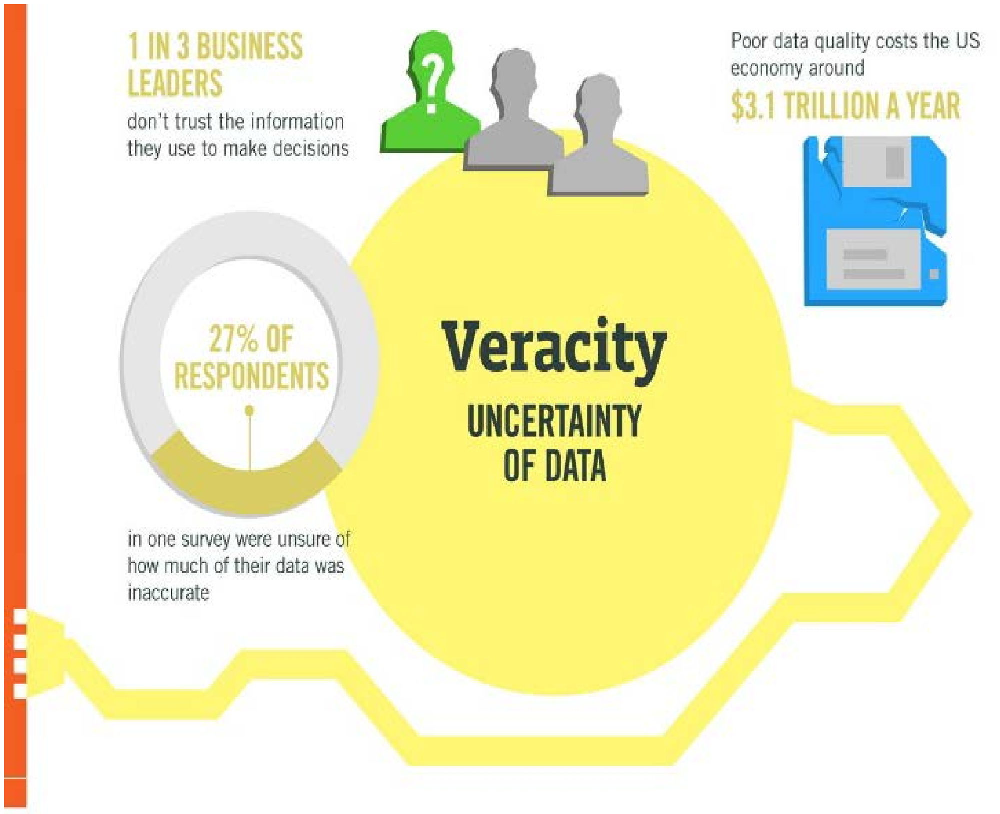
The Four Dimensions of Big Data: Velocity

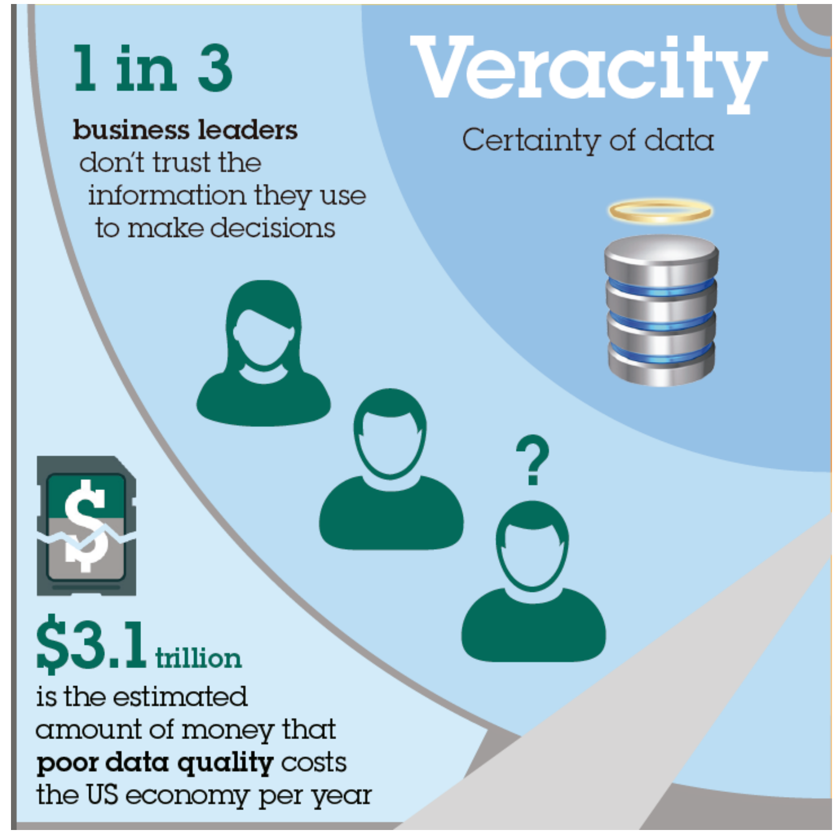




Source: www.ibmbigdatahub.com/infographic

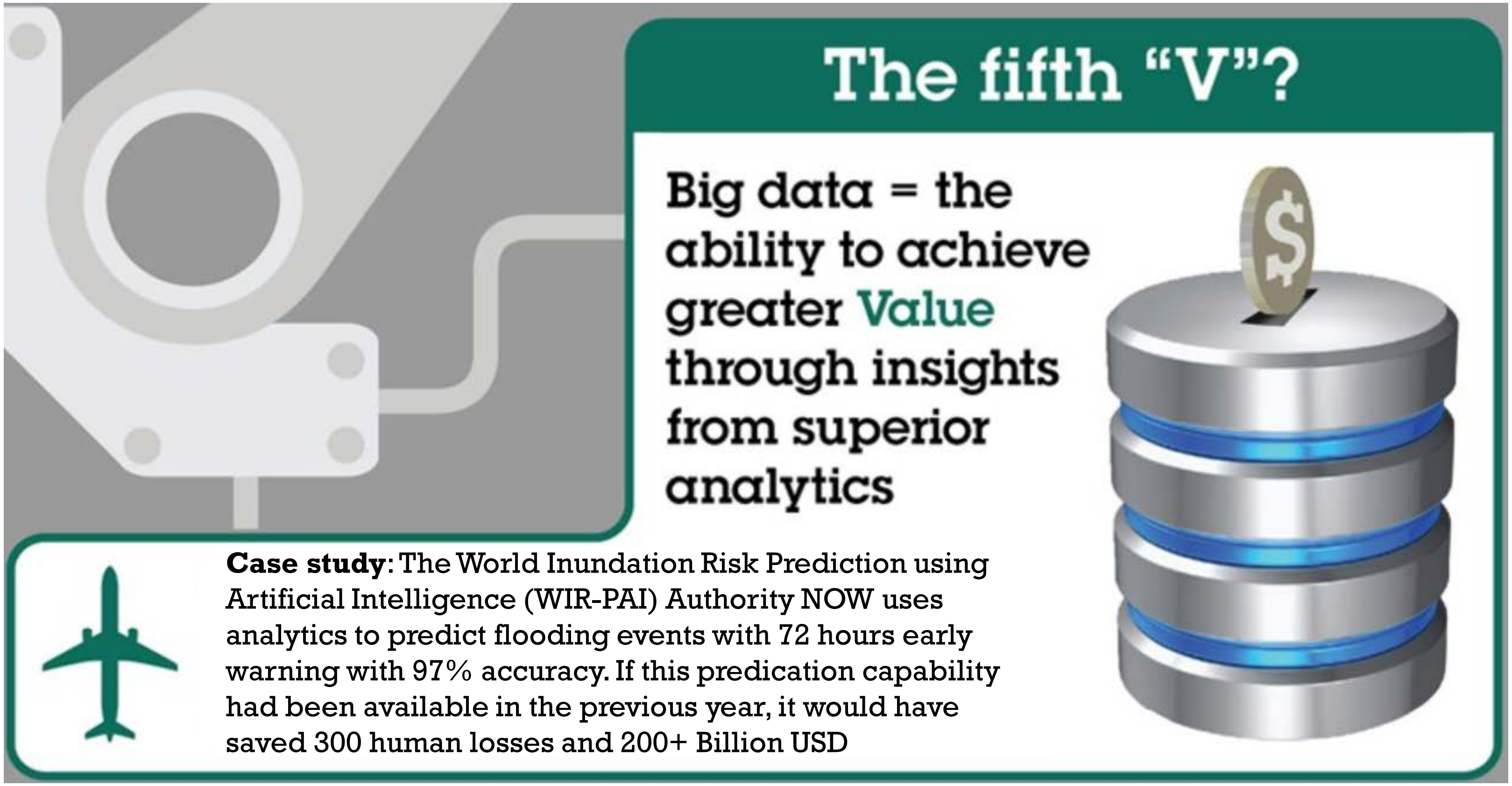
The Four Dimensions of Big Data: Veracity





Source: www.ibmbigdatahub.com/infographic

The Five Dimension of Big Data?



Source: www.ibmbigdatahub.com/infographic

Citizen Science in hydrology

The participation of the general public in the research design, data collection and interpretation process together with scientists is often referred to as citizen science.

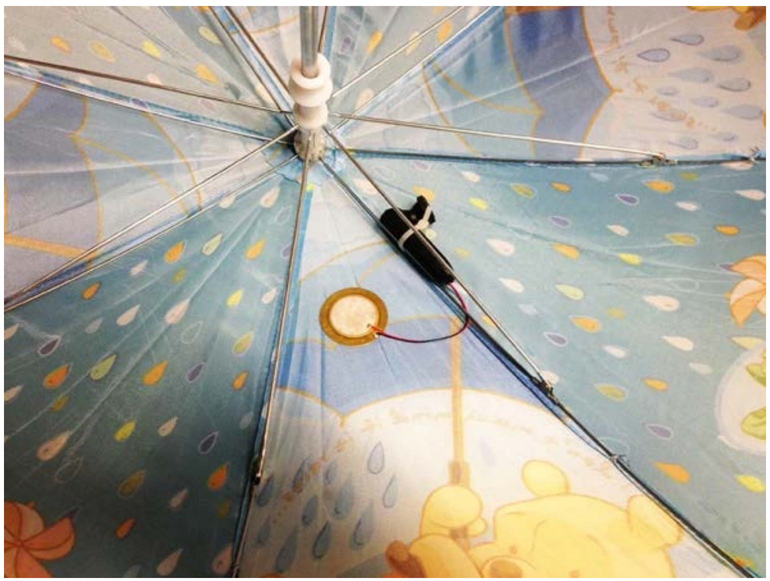
Variable	Opportunities	Challenges
Precipitation	Cheaper equipment (e.g., electronic tipping bucket rain gauges, disdrometers). Bulk analysis of environmental influences on rain captation. Merging with remotely sensed observations.	Proper installation, maintenance, and documentation of local environmental conditions. Long-term data collection.
Streamflow	Cheap and robust water level measurements; collection of calibration data; emerging image analysis techniques for stage and flow measurements.	Proper installation and maintenance; quality control; technical support.
Water quality	Cheap analysis toolkits; automatic measurement of proxies; macroinvertebrate observation and identification.	Several parameters remain costly and difficult to analyse; need for adequate documentation of observation context; sampling strategy.
Soil moisture	Automatic measurements (e.g., TDR) becoming increasingly affordable.	Relation with other soil properties; high spatial variability.
Vegetation dynamics	Very accessible technology (e.g., GPS, photography); remote identification.	Systematization; data processing; combination with remotely sensed data.
Water use	Availability of electronic sensors; convenient data communication via internet in built environments.	Interpretation and extrapolation of generated data; potential human interference.

Buytaert, et al. M. (2014). Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development. *Frontiers in Earth Science*, *2*, 26.

Citizen Science in hydrology

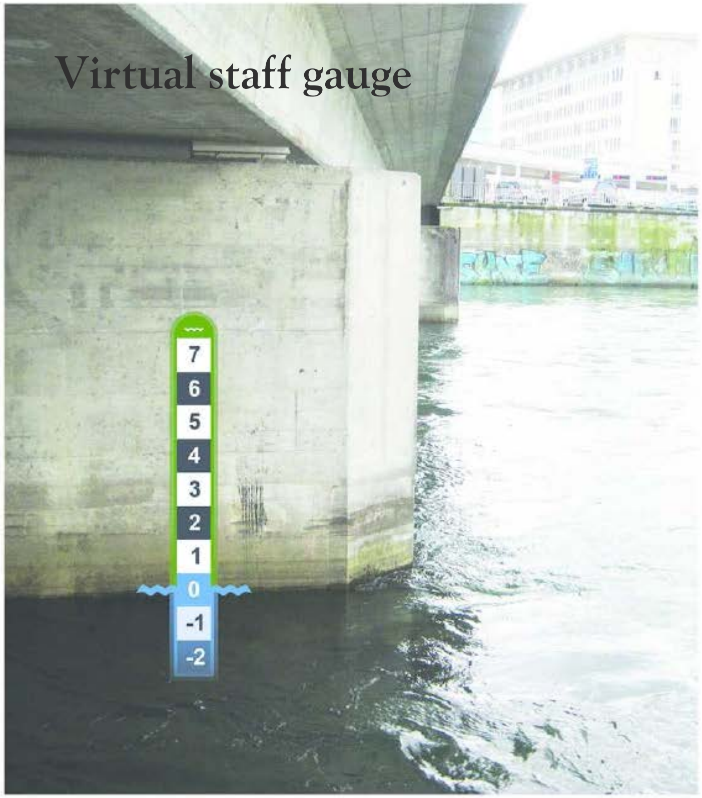
Measuring rainfall with umbrellas and car windshield wipers

Apps Wezzoo, Weddar and WeatherSignal all use data from mobile phone users to crowdshare local weather updates. **Many apps are (being) developed for crowdsourcing hydrologic data**



Hut, R., de Jong, S., & van de Giesen, N. (2014, May). Using umbrellas as mobile rain gauges: prototype demonstration. In EGU General Assembly Conference Abstracts (Vol. 16).
Rabiei, E., Haberlandt, U., Sester, M., & Fitzner, D. (2013). Rainfall estimation using moving cars as rain gauges-laboratory experiments. *Hydrology and Earth System Sciences 17 (2013), Nr. 11*, *17*(11), 4701-4712.

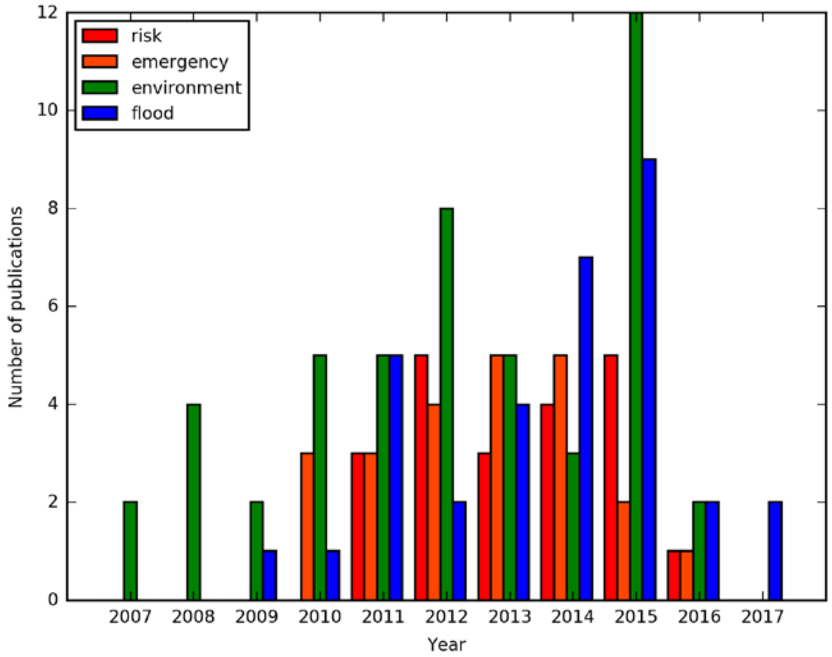
Citizen Science

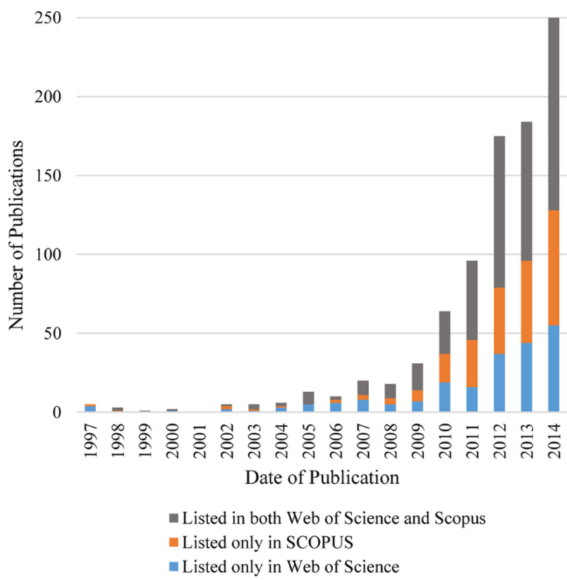


Check https://crowdwater.ch

Strobl Barbara, Etter Simon, van Meerveld Iljaa and Seibert Jan, Accuracy of Crowdsourced Streamflow and Stream Level Class Estimates, *Hydrological Sciences Journal* (Accepted)

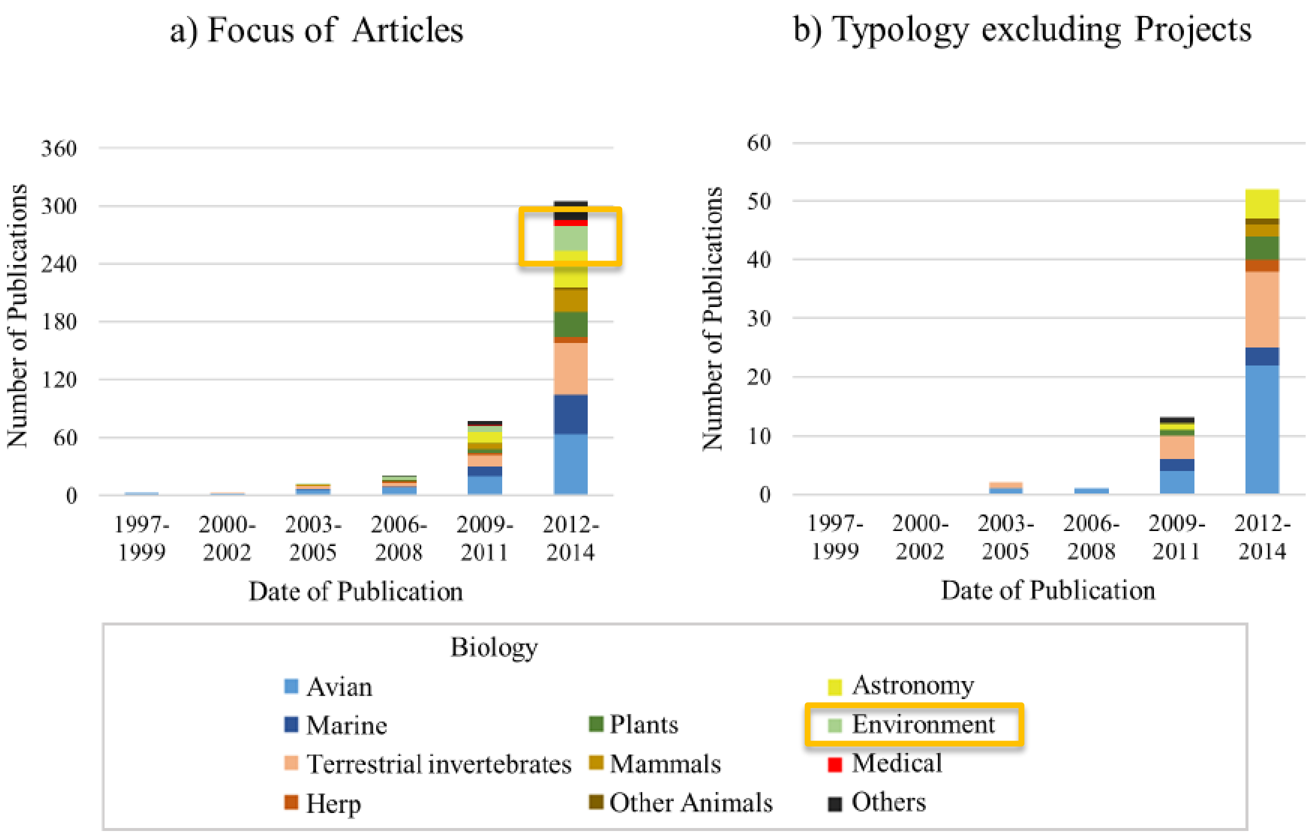
Citizen science data and «projects»





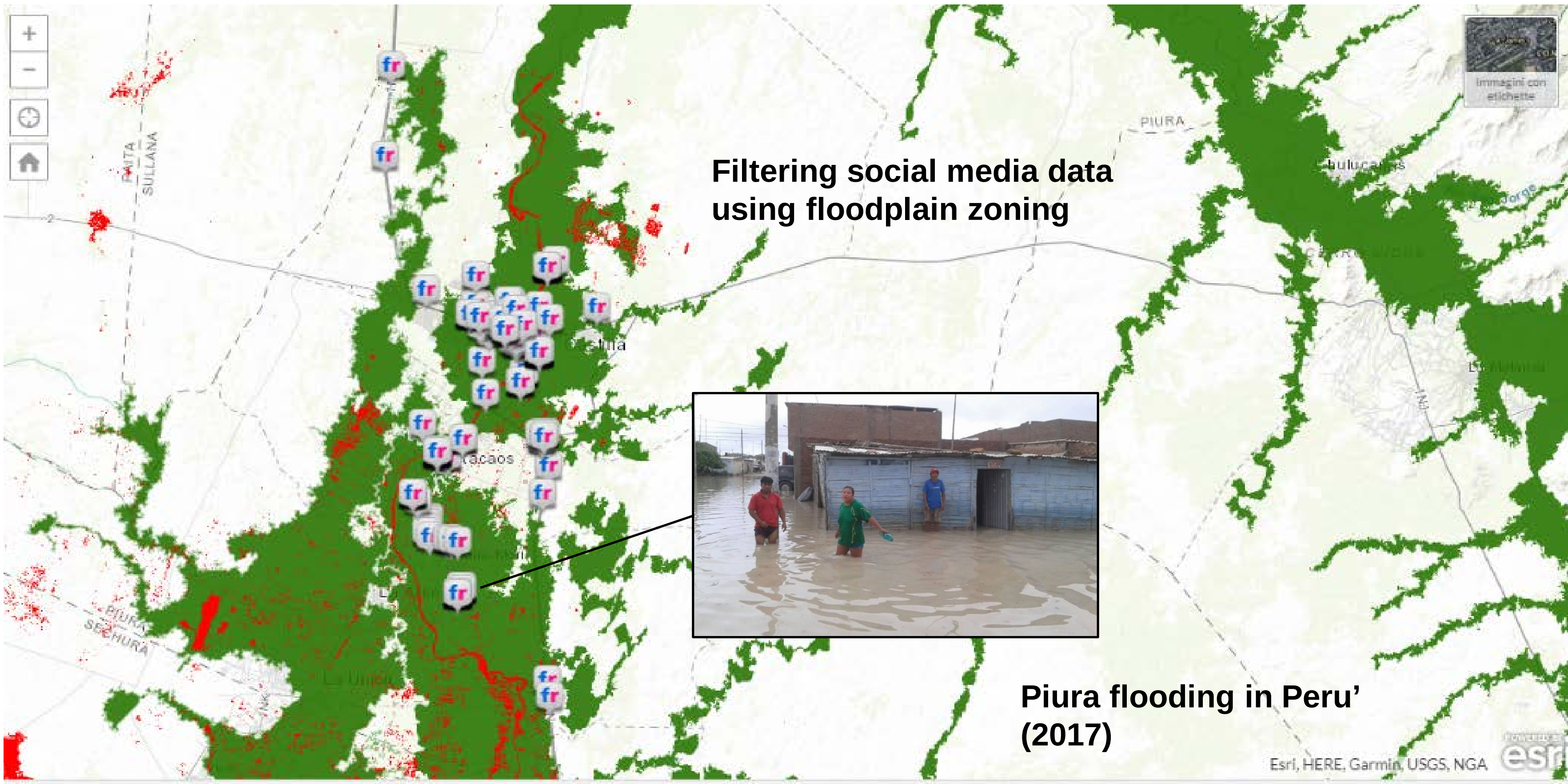
Follett & Strezov, 2015

Citizen science data and «projects»



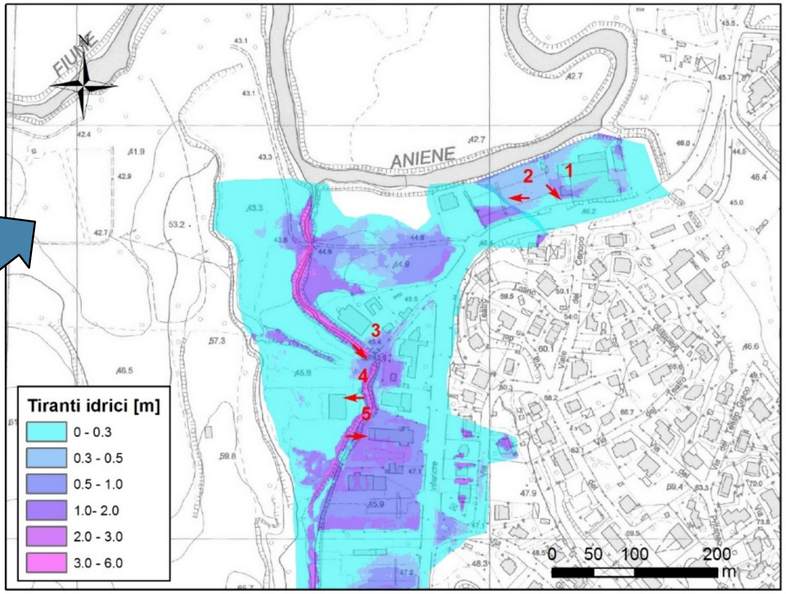
Follett, R., & Strezov, V. (2015). An analysis of citizen science based research: usage and publication patterns. *PloS one*, *10*(11), e0143687.

Human behavior/impact and geospatial mapping



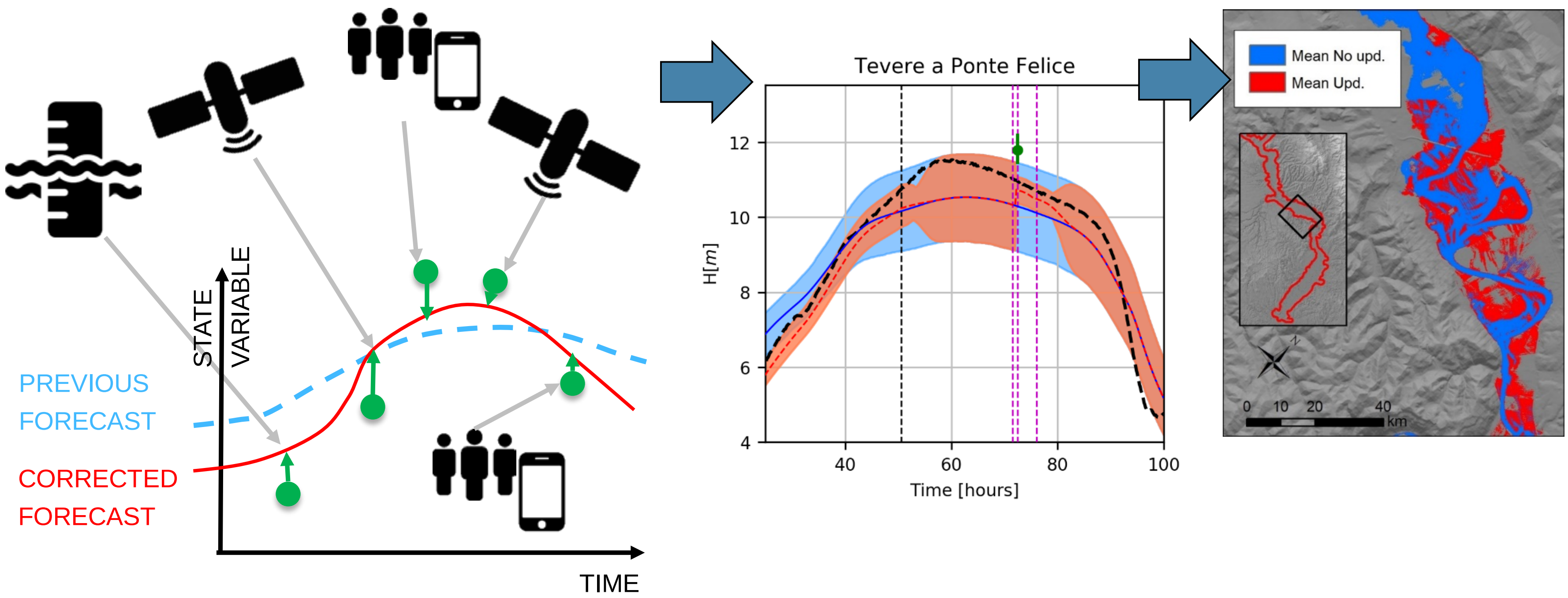
https://arcg.is/1umWbz

Human behavior/impact and geospatial mapping



Flooding & VGI - Rome San Vittorino flood, 2014
(in preparation)

Human behavior/impact/data and geospatial mapping



Data assimilation framework integrating diverse/several data sources (Antonio Annis et al., in preparation)

POP-Rain: citizen observations of spatial rainfall intensity

KEY POINTS

- Rainfall intensity can be qualitatively observed by citizens
- Just visually, citizens can provide useful information to hydrologists and civil protection
- Proper management of volunteers can improve citizen observations

S. Grimaldi, A. Petroselli, F. Tauro, A. M. Braccini, M. Porfiri, S. Noto, A. Cavalli & F. Nardi (team, manuscript in preparation)

POP-Rain

POP-Rain (i.e. POPular Rain) is a citizen science project that aims to verify if citizens can provide qualitative information on rainfall

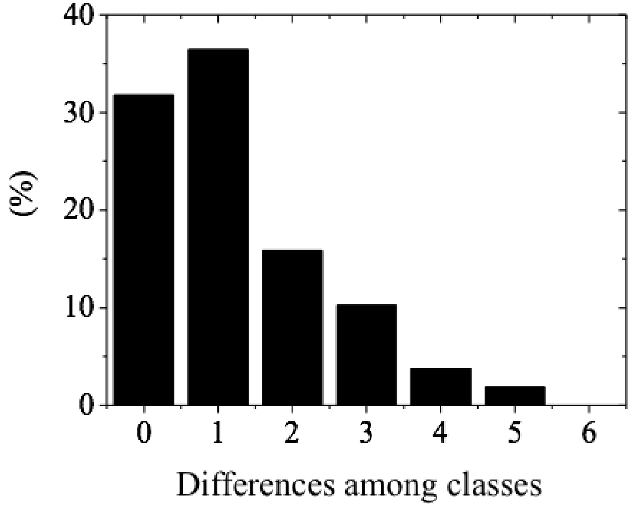
1) citizens are invited to quantify rainfall intensity just by looking at the natural phenomenon.
2) POP-Rain volunteers indicate their perception of rainfall intensity through a mobile app (rainfall intensity classes)
3) By collecting citizens' feedback, the POP-Rain app provides the citizens' geographical position and observations in real time.

CURRENT STAGE

- 150 volunteers recruited among the student community at Tuscia University of Viterbo
- since 2017, more than 300 observations
- validation made by comparing the qualitative evaluation provided by volunteers to actual rainfall measurements observed at the official rain-gauge stations located closest to POP-Rain observers

PRELIMINARY RESULTS

- more than 30% of the observations were in agreement with the raingauge (bar 0)
- more than 35% of the observations showed an error of only one class (bar 1).
- Almost 70% of the observations were correct or reasonable.
- 10% of the observations were totally misinterpreted



1) Pioviggine (< 1mm/h)
2) Pioggia debole (1-2 mm/h)
3) Pioggia leggera (2-4 mm/h)
4) Pioggia moderata (4-6 mm/h)
5) Pioggia forte (>6 mm/h)
6) Rovescio (> 10 mm/h)
7) Nubifragio (> 30 mm/h)

POP-Rain APP

- Android app free to download and use
- Easy to use interface
- Wizard style process
- User access with account/password protected
- Proper naming and definitions of key words and parameters
- Geotagged and GPS-supported information platform
- Easy database export for project managers

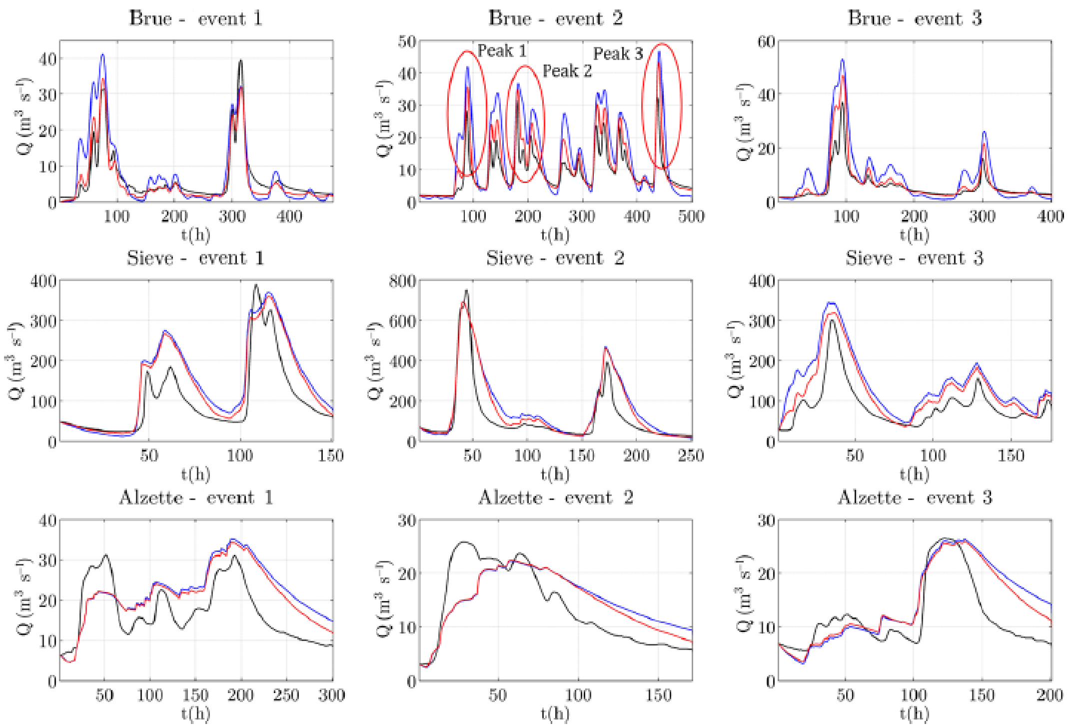


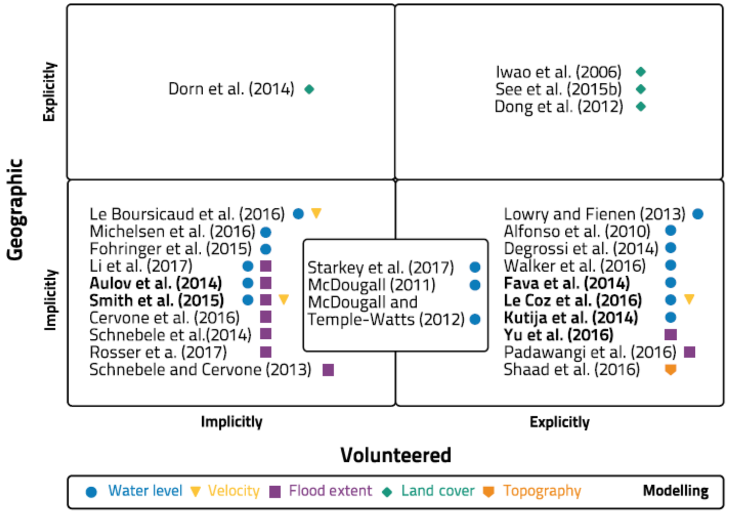
Figure 7. Observed (black line) and simulated hydrographs, with (red line) and without (blue line) assimilation, for the flood events which occurred in the three catchments: Brue (upper row), Sieve (middle row), and Alzette (bottom row).

Data assimilation
using Citizen science for
improved flood hazard management

Mazzoleni, et al. (2017). Towards assimilation of crowdsourced observations for different levels of citizen engagement: the flood event of 2013 in the Bacchiglione catchment. Hydrology and earth system sciences.

Mazzoleni, et al. (2017). Can assimilation of crowdsourced data in hydrological modelling improve flood prediction? Hydrology & Earth System Sciences, 21(2).

Citizen science data/tools for hydrologic applications



studies represented in the typology of VGI (volunteered geographic information).

Review paper

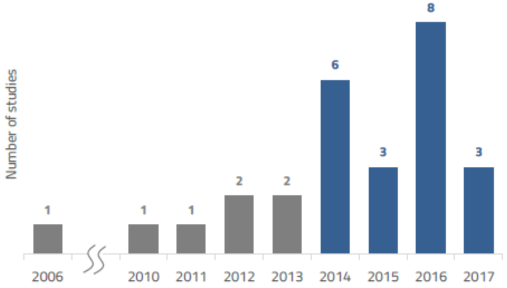


Figure 4. Number of studies analysed per year.

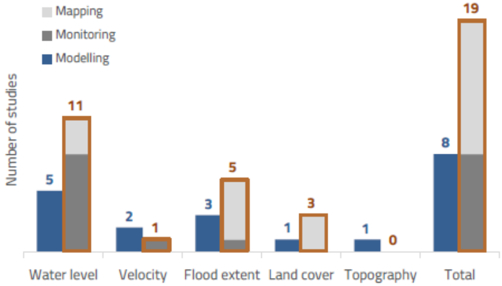
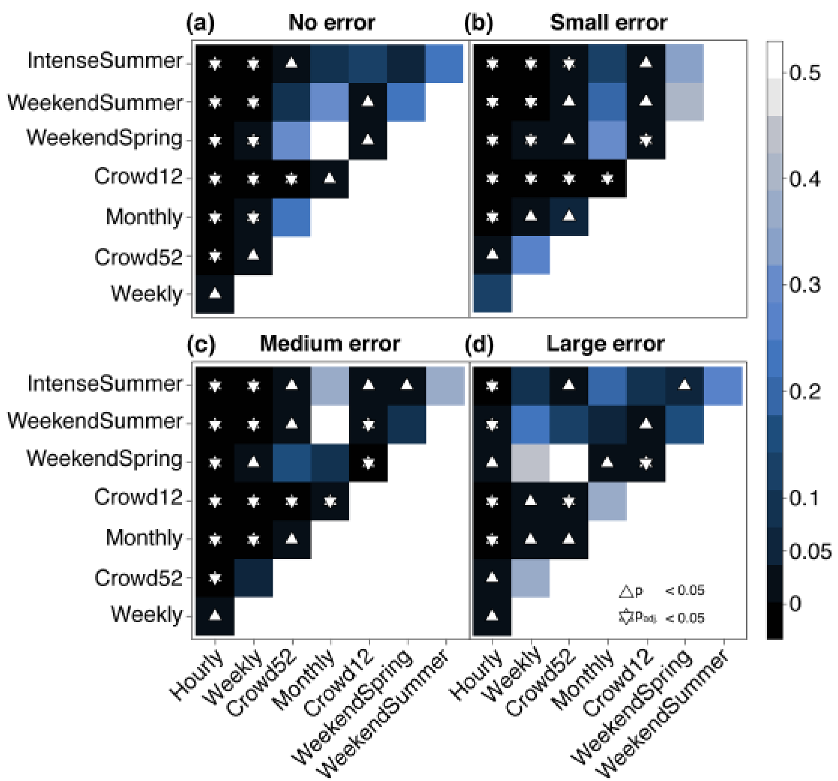


Figure 5. Number of studies analysed per flood-related variable per category: mapping, monitoring and modelling.

Assumpcao, T. H., I. Popescu, A. Jonoski, and D. P. Solomatine (2018). Citizen observations contributing to food modelling: opportunities and challenges. Hydrology and Earth System Sciences 22 (2), 1473-1489.

Citizen science data/tools for hydrologic applications



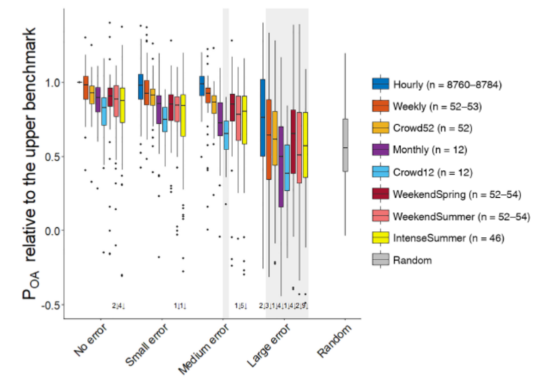


Figure 4. Box plots of the median model performance relative to the upper benchmark for all datasets. The grey rectangles around the boxes indicate non-significant differences in median model performance compared to the lower benchmark with random parameter sets. The box represents the 25th and 75th percentile, the thick horizontal line represents the median, the whiskers extend to 1.5 times the interquartile range below the 25th percentile and above the 75th percentile and the dots represent the outliers. The numbers at the bottom indicate the number of outliers beyond the figure margins; *n* is the number of streamflow observations used for model calibration. The result of the hourly benchmark FOEN dataset has some spread because the results of the 100 parameters sets were divided by their median performance. A relative P_{OA} of 1 indicates that the model performance is as good as the performance of the model calibrated with the hourly FOEN data (upper benchmark).

Etter, S., Strobl, B., Seibert, J., & Meerveld, H. J. (2018). Value of uncertain streamflow observations for hydrological modelling. *Hydrology and Earth System Sciences*, *22*(10), 5243-5257.

Citizen science data/tools for hydrologic applications

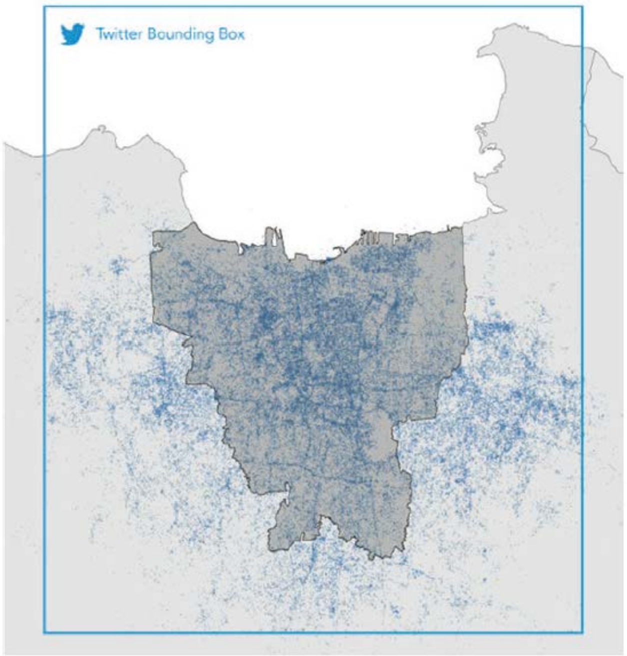


Fig. 2 Map of tweets related to flooding during the 2013–2014 monsoon season. *Data source* Twitter #DataGrant; [11]

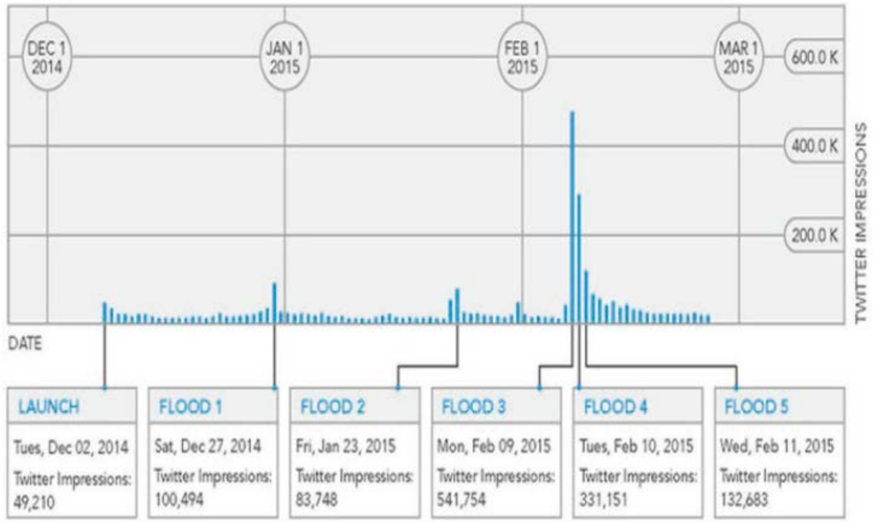


Fig. 3 Timeseries of Twitter impressions from the *@petajkt* account during five flood events of the 2014–2015 monsoon season. *Data source* Twitter Analytics; [11]

Holderness, T., & Turpin, E. (2015). From social media to geosocial intelligence: Crowdsourcing civic co-management for flood response in Jakarta, Indonesia. In *Social media for government services* (pp. 115-133). Springer, Cham.

Citizen science data/tools for hydrologic applications

Flickr semantic for flood monitoring/forecasting. No modelling, no mapping, filtering non structured data using flood-related keywords using social media

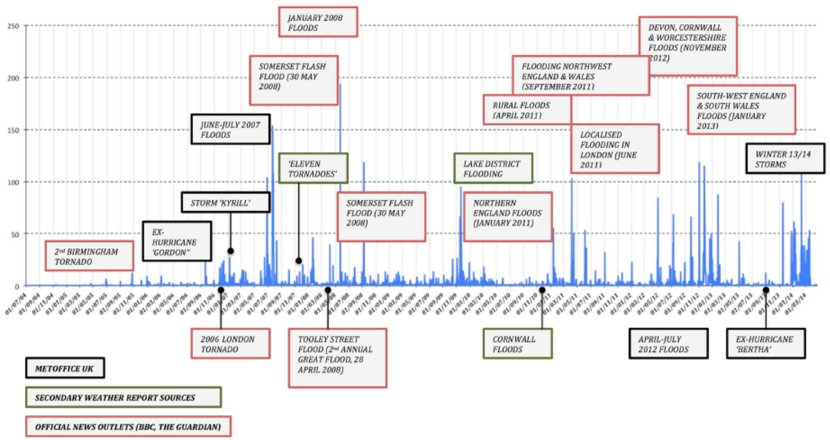


Fig 4. Timeline of Flickr activity during a number of flood events (2004–2014) of various magnitude, captured by official, secondary and mass media sources.

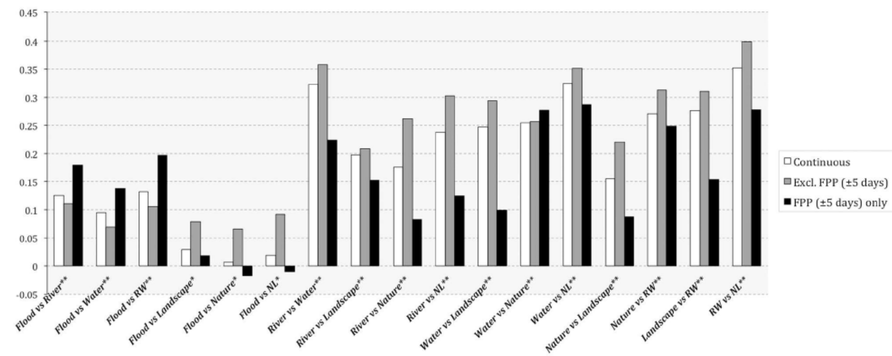
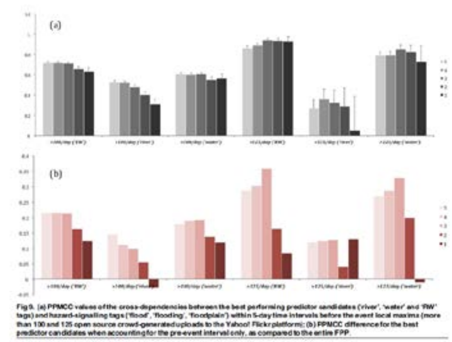


Fig 6. PPMCC values of the cross-dependencies between generic environmental, hydrological and risk-signalling semantic tags on the Yahoo! Flickr platform (2004–2014).



Tkachenko, N., Jarvis, S., & Procter, R. (2017). Predicting floods with Flickr tags. *PloS one*, *12*(2), e0172870.

Citizen science data/tools for hydrologic applications

Big Data and Computer vision (Artifical intelligence) using social media for flood monitoring

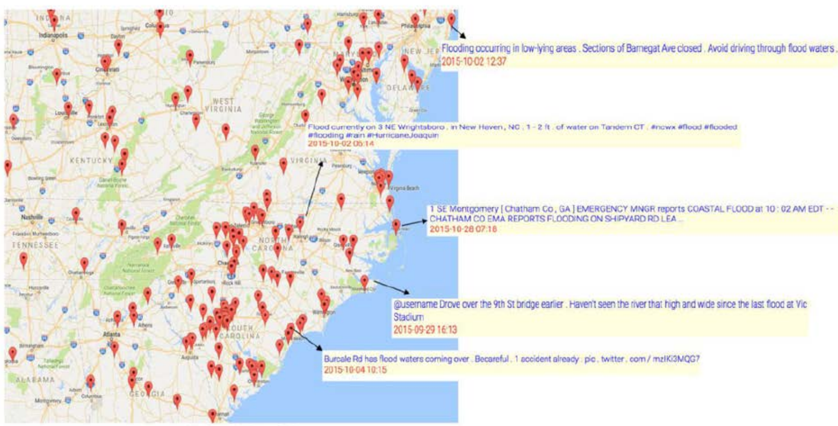


Fig. 4. The spatial distribution of identified Tweets that have hyper-resolution geo-location information. Text content of sample posts is shown in the right panel.

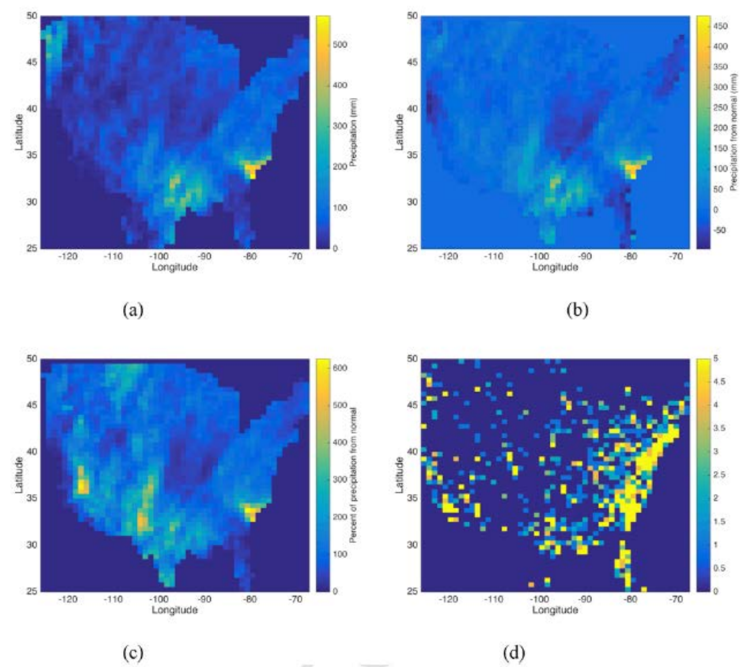
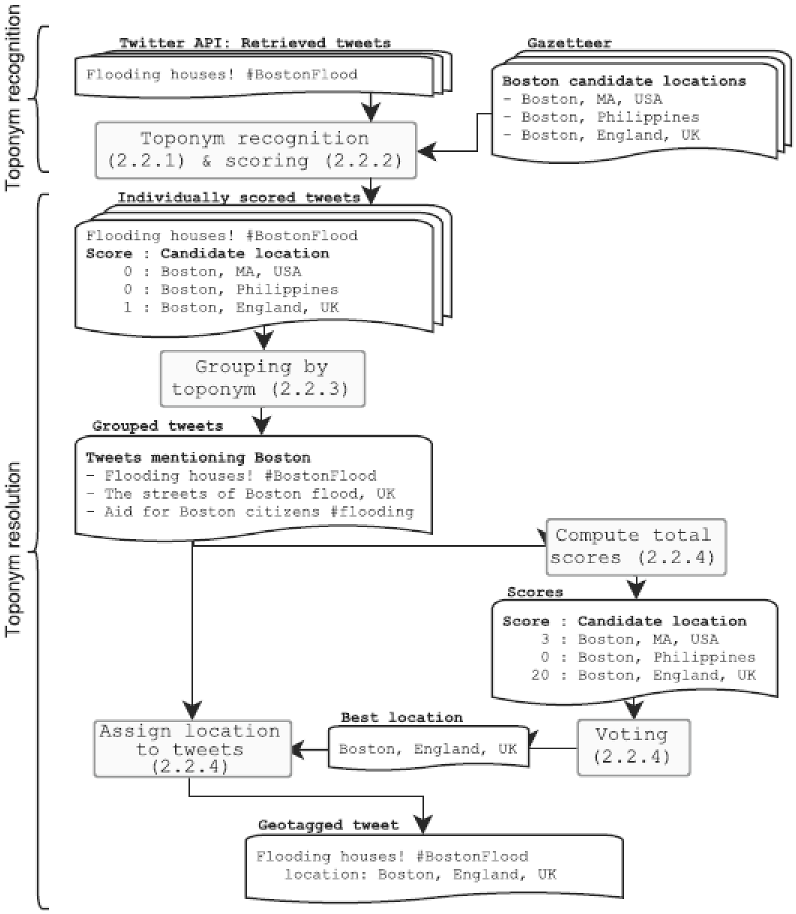


Fig. 5. Spatial patterns of precipitation and Tweets in October, 2015. (a) The observed precipitation. (b) The precipitation departure. (c) The percentage of precipitation departure. (d) The statistics of the Tweet volume.

Wang, R. Q., Mao, H., Wang, Y., Rae, C., & Shaw, W. (2018). Hyper-resolution monitoring of urban flooding with social media and crowdsourcing data. *Computers & Geosciences*, *111*, 139-147.

Fig. 2 Overview of the TAGGS geoparsing process



Flood detection using social media tags

de Bruijn, J. A., de Moel, H., Jongman, B., Wagemaker, J., & Aerts, J. C. (2018). TAGGS: Grouping tweets to improve global geoparsing for disaster response. *Journal of Geovisualization and Spatial Analysis*, *2*(1), 2.

International Association of Hydrological Sciences (IAHS)

IAHS Working Group on on “Citizen and Hydrology” (CANDHY)

https://iahs.info/Commissions--W-Groups/Working-Groups/

Working Groups
Panta Rhei (everything flows)
Education in Hydrological Sciences
MOXXI – Measurements and Observations in the XXI Century
CandHy (approved at the IAHS Bureau Meeting, Port Elizabeth, South Africa, July 2017)

IAHS AISH

CANDHY

CANDHY

The CandHy Working Group (WG) aims to stimulate discussion, sharing of knowledge, information, data, ideas fostering scientific and professional exchange of academic, institutional and citizen communities interested in the "Citizen AND HYdrology" topic. CandHy seeks to discover the potential of citizen involvement and crowdsourced data for advancing hydrologic research in water resource and risk management, hydroclimatic risk mitigation and disaster preparedness. CandHy addresses the challenges stated by IAHS for the scientific decade 2013–2022 entitled "Panta Rhei – Everything Flows" on hydrology and society with specific regard to the value of citizen science and crowdsourced data in studying the mutual feedbacks between natural and anthropogenic dynamics.

CandHy promotes multidisciplinary and trans-sectorial researches integrating hydrology, and earth science in general, with citizen and data science. CandHy investigates on novel methods for adopting user generated content (UGC) (e.g. data from social network) and volunteering geographic information (VGI) for collaborative analysis and mapping (e.g. OpenStreetMap) with the aim of advancing in the understanding, analysis, modelling and communication of hydrologic dynamics. CandHy explores the role of citizens acting as human sensors of hydrologic and societal processes and features using largely accessible technology (e.g. smart phones, web cams, drones, …) and, at the same time, providing information on population dynamics, citizen perception and behavior that shall be used by decision makers for water risk management.

The focus of CandHy WG is to organize, share and disseminate ideas, tools, data and scientific knowledge gathered from multidisciplinary and trans-sectorial researches developing and testing innovative procedures and tools implementing crowdsourcing collaborative platforms, communication and outreach activities, active citizenship and public participation for water resource management.

CandHy aims to become an open hub where the WG community members, partners and citizens (CandHy friends) meet and explores the value of both citizen and expert volunteering for knowledge and data creation and sharing. Citizen involvement and the use of novel observational techniques are key themes for addressing hydrologic challenges and communicating water risk, especially in developing countries where the lack of monitoring and education can greatly benefit

EVENTS
CANDHY FRIENDS
PUBLICATIONS
PROJECTS AND DATA

WORKING GROUPS
CANDHY
PANTA RHEI
PREDICTION IN UNGAUGED BASINS
EDUCATION IN HYDROLOGICAL SCIENCES
MOXXI
ICCE, CONTINENTAL EROSION
ICCE NEWS
ICCE PUBLICATIONS
OBITUARY: PROFESSOR WAYNE ERSKINE
ICCE EVENTS
ICCLAS, COUPLED LAND

Citizen and Hydrology (CandHy) Working Group

https://iahs.info/Commissions--W-Groups/Working-Groups/Candhy.do

Iahs.info -> Commissions -> Working Groups -> CANDHY

Goals

- Stimulate discussion, sharing of knowledge, information, data, ideas
- Fostering scientific and professional exchange of academic, institutional and citizen communities
- Discover the potential of citizen involvement and crowdsourced data for advancing hydrology
- Interact and interlink with other working CS-related platforms and working groups (Panta Rhei, MOXXI from IAHS, and others worldwide)

Citizen and Hydrology (CandHy) Working Group

https://iahs.info/Commissions--W-Groups/Working-Groups/Candhy.do

Iahs.info -> Commissions -> Working Groups -> CANDHY

Projects

- Archiving/Linking CandHy data (sources), projects, tools, models, etc
- Identify and share journals, papers and publications/reports
- Spread the word: mailing/networking list, website, newsletter, conferences
- Supporting the involvement of research groups and young scientists/students (social networks, IAHS Linkedin group, …)

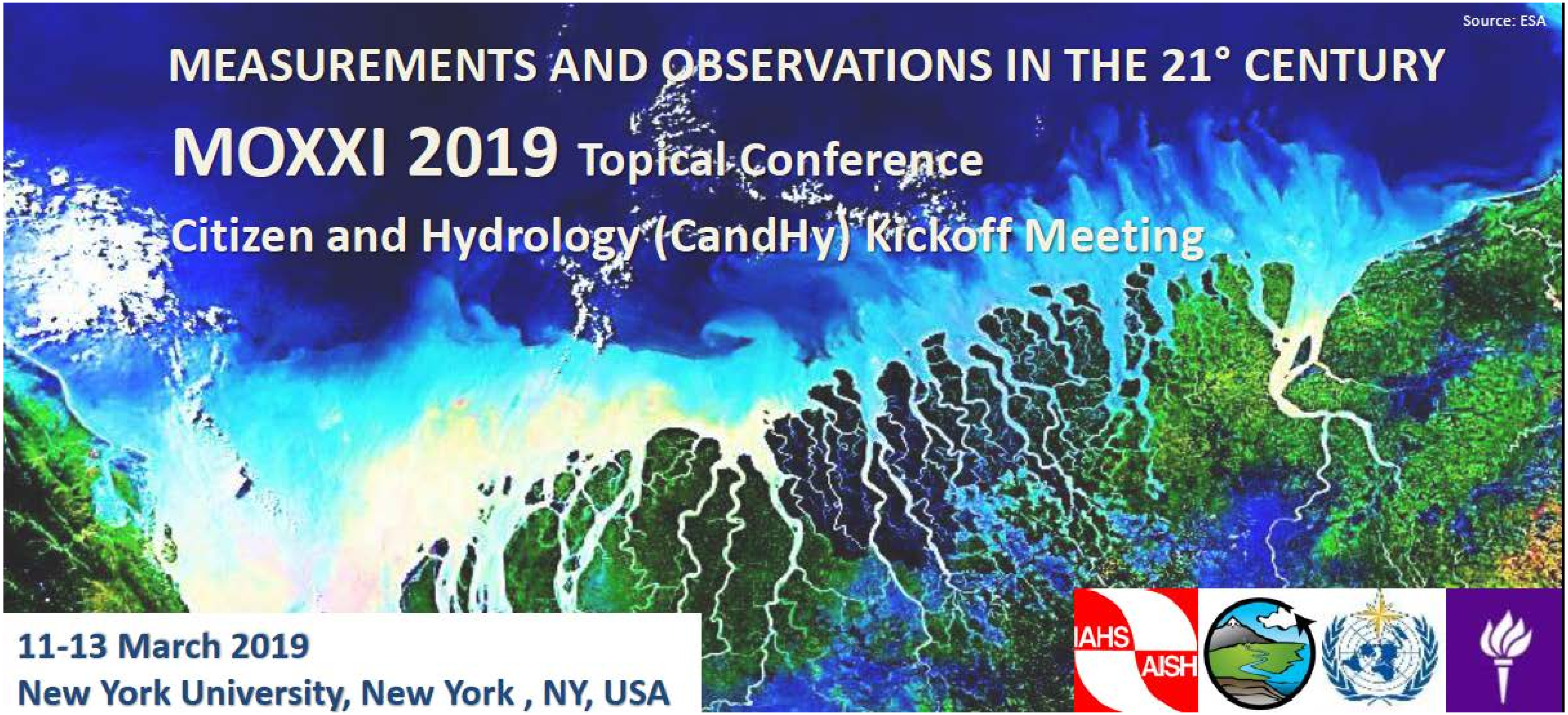
Events

- April 11, 2018 CandHy IceBreaker event at EGU General Assembly 2018 & Kick off the CandHy friends (Vienna, Austria)
- November 22, 2018 International Workshop: Open/Big Data and Citizen Science for Managing the Water, Food, Energy and Environment Nexus (Perugia, Italy) (http://warredoc.unistrapg.it/ Go to events section)
- November 27, 2018. SMIRES Cost Action "Citizen Science Workshop"
- November 28-29, 2018. COWM Conference (Venice, Italy)

Coming next…

- 2019 Moxxi, Candhy, Wmo Hydrohub, & Cuahsi Joint Conference "Innovation In Hydrometry: Overcoming Barriers To Operationalization" – New York, USA (March 11-13, 2019)
- IUGG General Assembly (Montreal, Canada – July 8-18, 2018)

Call for abstracts & Special Issues



Call for abstracts & Special Issues

- Hydrological Sciences Journal Special Issue on "**Hydrological Data: Opportunities and Barriers**" (deadline: 31st March 2019)
- Virtual Special Issue: "**Advancing socio-hydrology: a synthesis of coupled human–water systems across disciplines**" (just launched on november 27, 2018)


...snowing at Villa Colombella...
..this is a WINTER School!!

CINID CONSORZIO INTERUNIVERSITARIO PER L'IDROLOGIA | UNESCO United Nations Educational, Scientific and Cultural Organization | World Water Assessment Programme | WARREDOC | Università per Stranieri di Perugia

2019 Edition Doctoral Winter School DATA RICH HYDROLOGY

Winter School RUN

- Tomorrow morning (feb 1st)
- 8k run/walk/sleep/______along the____Tiber river
- 7am SHARP!

..raising water awareness, team building,
stay healthy, observing hydrologic processes at work!!

Stefan Uhlenbrook @SUhlenbrook Following

We started Day II of the Winter School by @UNISTRAPG @warredoc and #UNWWAP on 'Data Rich Hydrology' with a morning run with snowfall in Perugia (Italy) and we would like to donate our kilometres to #minaguli and #RunningDry

6 × 8 km = 40 km

@fnardi

12:22 - 29 gen 2019

You can't miss it!

Bibliografia

- Bahrami, M., Singhal, M., (2014). The Role of Cloud Computing Architecture in Big Data. vol. 8. Springer International Publishing, Cham, pp. 275 295. http://dx.doi.org/10.1007/978-3-319-08254-7_13.
- Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, *130*, 99-113.
- Cox, M., Ellsworth, D., (1997). Managing Big Data for scientific visualization. ACM
- Russell, M.A., (2013). Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub and More, 2nd. O'Reilly Media, pp. 448.
- Wessels, B., Finn, R., Wadhwa, K., & Sveinsdottir, T. (2017). Open data and the knowledge society. Amsterdam University Press.
- Xiong, T., Wang, S., Mayers, A., Monga, E., (2013). Personal bankruptcy prediction by mining credit card data. Expert Syst. Appl. 40, 665–676. http://dx.doi.org/10.1016/ j.eswa.2012.07.072.

Some selected Glossary resources

https://www.zentrumfuercitizenscience.at/en/glossary
http://opendatahandbook.org/glossary/it/
https://citizenscienceguide.com/glossary

Thanks for the attention

http://warredoc.unistrapg.it

fernando.nardi@unistrapg.it