Next-generation geospatial-temporal information technologies for disaster management

C. M. Albrecht
B. Elmegreen
O. Gunawan
H. F. Hamann
L. J. Klein
S. Lu
F. Mariano
C. Siebenschuh
J. Schmude

Traditional geographic information systems (GIS) have been disrupted by the emergence of Big Data in the form of geo-coded raster, vector, and time-series Internet-of-Things data. This article discusses the application of new scalable technologies that go far beyond relational databases and file-based storage on spinning disk or tape to incorporate both storage and processing data in the same platform. The roles of the Apache Hadoop Distributed File Systems and NoSQL key-value stores such as the Apache Hbase are discussed, along with indexing schemes that optimally support geospatial-temporal use. We highlight how this new approach can rapidly search multiple GIS data layers to obtain insights in the context of early warning, impact evaluation, response, and recovery to earthquake and wildfire disasters.

Introduction

Geographic information systems (GIS) have been around for a long time. One of the earliest examples of geographic analysis is the well-known work of John Snow in tracing the location and source of a cholera outbreak in London in 1849 [1]. The term "GIS" was first coined by Roger Tomlinson in 1968, who was at the center of a Canadian land mapping application [2]. The traditional data formats used in GIS are vector and raster imagery with sparse temporal and limited spatial coverage, but the dropping cost of sensors and imagery acquisition combined with novel sensing system like geo-tagged mobile, social, and crowdsource data increased data volume exponentially. While a few years back, a couple of terabytes were generated yearly from earth observing satellites likes the land remote-sensing satellite system (Landsat) and the moderate resolution imaging spectroradiometer (Modis) imagery, currently GIS data generation is in the range of hundreds of petabytes per year [3]. The high density and wide coverage of earth-surface data enable researchers to discover and monitor changes remotely in near-real time.

With the extensive use of GIS data, new applications have been developed in many industries to assess "what and where" events are happening [4]. For example, in the telecom industry, GIS data can help determine where to build the next-generation cell tower network. In agriculture, it might show where to apply fertilizers or plant the best crop on a farm. In disasters like hurricanes, it can show government agencies where to send first response teams and what equipment is needed. Companies like Uber, Foursquare, Zillow, and Garmin have merged GIS data with social data and created new businesses [5] built on open source data. The diversity and the volume of data generated require a paradigm shift in data storage and processing.

Traditional GIS data stores, which are generally using a relational database with spatial extender in back end, are being disrupted by the emergence of Big Data technologies that offer a more comprehensive data management. First and foremost, the nature of relevant GIS data is changing drastically. Traditional GIS is designed to work with mostly static spatial vector data, which is generally small-volume. For example, all the vector data from openstreetmap.org, which includes most roads and houses across the globe as well as increasing coverage of human infrastructure such as houses, stores, land use, traffic signs, hiking trails, etc., is only a few gigabytes in size. There is much more data—often orders of magnitude more—that is in the form of geo-coded imagery from drones, cell phones, and satellites. The data generation rate of the European Space Agency (ESA)

Digital Object Identifier: 10.1147/JRD.2020.2970903

(c) IBM 2020. This article is free to access and download, along with rights for full text and data mining, re-use and analysis.

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 01,2024 at 03:25:00 UTC from IEEE Xplore. Restrictions apply.

is tens of terabytes per day. Point cloud vector data from Light Detection and Ranging (LIDAR), which consists of billions of reflected intensities from ground-pointing lasers, has approximately 10 to 30 data points per square meter for high-resolution mapping. Finally, the Internet of Things (IoT) produces enormous volumes of geo-coded temporal, real-time data. For example, geo-coded IoT data from the advanced metering infrastructure monitors energy consumption at each endpoint in an electric power grid. Overlays of such real-time energy consumption data with other data such as weather, renewable energy generation, etc., would enable more efficient grid operations and faster restoration of power after a disaster. Traditional GIS was not primarily designed for real-time data and is aimed at "what and where" use cases leveraging small-volume vector data. However, today many "what, when, and where" use cases have emerged, and these involve much larger volumes of data of various types (vector and raster), including the time dimension (static and temporal). The next disrupting trend can be described by the notion of "data gravity," which is a result of two facts. First, geospatial-temporal data is becoming so big for many use cases that it cannot be downloaded to a local computer or moved. The 10 TB of data generated daily by ESA would take more than a day to download to a single disk drive. Second, geospatial-temporal data, like other data, can be tremendously enhanced by other contextual information that is linked in space and time. These links allow multiple datasets to be overlaid to identify meaningful statistical dependencies or enhance the fidelity of machine-learning models. On an engineering level, data gravity means that the next-generation GIS must heavily leverage cloud computing to reach scalability. The additional need for private data must be addressed using hybrid implementations. In addition, and equally important, is the requirement of new architectures and systems where the computation is "moved" to the data rather than the traditional way, where data are transferred to some local memory near a processor. Only with highly parallel processing close to the data can this massive and rich source of geospatial-temporal information be fully exploited in a timely matter.

As a consequence of data gravity, and to reduce the amount of data movement, it has become clear that geospatial information, especially imagery, has to be indexed on a much more refined granularity than before so that analytical tasks can be performed within the query or data request itself, as opposed to by the application. Today, most geo-encoded raster data are only indexed at the level of the metadata associated with it, but what if someone wants to search for certain data *within* the raster dataset? A simple example would be to prepare an emergency response to a major storm by searching weather forecast maps, population distributions, and elevation data. A realistic query might be: When and where will there be more than 10 inches of rain in the next 24 hours, where the elevation is lower than sea level and the population density is larger than 50 people per square km?

Today, such a search is performed after the various datasets are downloaded or moved to the processors running the user's application. The requirement for finer indexing is also important because many geospatial-temporal datasets include physical quantities that one might want to use directly. For example, weather datasets might include temperature, precipitation, and wind speed in large files, and only the temperatures are needed. Moreover, satellite images might have several spectral bands that can be combined for new insight.

As the amount of data acquired and stored in satellite imagery is increasing exponentially, the ability to run analytics and extract information from the data requires new data storage architecture and coupling analytics with the data. The current trends in geospatial data management systems are described along with an implementation of a Big Data platform called PAIRS Geoscope [6] that enable large-scale analytics for disaster management. The performance of PAIRS Geoscope is compared to conventional file store platforms based on a set of queries that are relevant demonstrating the advantages of spatially and temporally aligned datasets for disaster management scenarios.

Next-generation GIS–IBM PAIRS geoscope

In the following, we describe an instantiation of a next-generation GIS, the PAIRS Geoscope, which goes much beyond current (commercial) implementations using relational databases or files/object stores [6, 7]. PAIRS, an acronym for Physical Analytics Integrated Data and Repository Services, is specifically crafted for the complexity and size of geospatial-temporal information.

As shown in **Figure 1**, PAIRS Geoscope entails three components: Built on top of GDAL/OGR (an open-source Geospatial Data Abstraction Library [8]), there is an *ingestion or data curation engine* to support the upload of more than 200 different formats (geotiff, grib, geojson, etc.) that can seamlessly integrate all forms of geospatial-temporal data, including weather data and models, satellite and aerial imagery (raster data), map-like information (vector data), and data from georeferenced devices and sensors, including social and event data (e.g., Twitter and GDELT) in standard or customized formats. During ingestion, all data are curated by validating the quality, aligning to a global reference system, and reprojecting or resampling. The current parallel ingestion engine is based on optimized C++ routines that allow to process over 20 TB per day. Future versions will enhance the ingestion engine by adding computational resources and expanding its capabilities through orthorectification, three-dimensional (3-D) mapping, and augmented data quality control.

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 01,2024 at 03:25:00 UTC from IEEE Xplore. Restrictions apply.

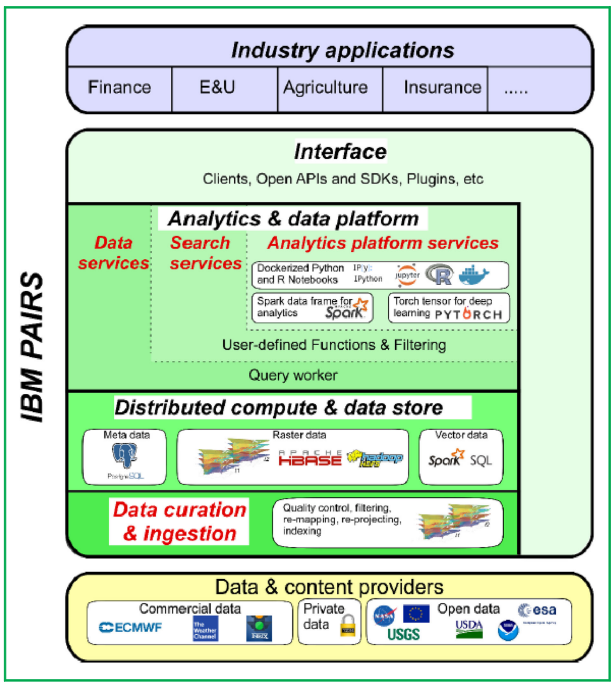

Figure 1

Overview of PAIRS Geoscope architecture.

Once ingested to the *core cluster*, large, heterogeneous, and complex datasets are tamed into a tidy aligned "layered" structure. The data layers are stored as multilevel key-value combinations. The key index encodes space, time, and other dimensions required to uniquely represent the value (e.g., for weather forecast, Δt, the forecasting time horizon), while the value represents data from a single pixel or an array of pixels (a "supercell"). The keys and values are indexed for efficient retrieval, conditional filtering, and aggregation. This core cluster is built on an open-source distributed file system and key-value technology (Hadoop/HBase) that is scalable up to hundreds of petabytes. Users can upload their own data to securely segregated "private" layers while still having access to all "public" layers. The PAIRS Geoscope system is being extended for big vector data by leveraging other technologies (Geomesa and/or Geowave, which are open-source, distributed, geospatial-temporal indexing systems built for Bigtable-style databases [9, 10]) including point cloud data processing. New data management libraries like Point Data Abstraction Library complement the standard GDAL functions to quickly cross the boundary from raster-based data to 3-D object recognition and classification.

RESTful APIs and SDKs (for Python, R, etc.) enable users to *interface* with the system, whether this includes the PAIRS Geoscope client (an interactive tool to create queries) or any other third-party or custom software. IBM has open-sourced a Python API wrapper: https://github.com/IBM/ibmpairs. Applications can be built on top of PAIRS. The results from a query are available—to avoid moving the data—as an in-memory data frame for further analytics, or they can be converted to different formats (ESRI Shapefile, GeoJSON, GeoTiff, Comma Separated Value text file) for downloading.

The first distinguishing characteristic of PAIRS to the best of our knowledge is being the only "commercial" big geospatial-temporal platform that indexes raster data on a pixel or supercell level. The basic technical challenges lie in the design of the key and the supercell as well as efficient indexing. Today, commercial big geospatial-temporal data analytics platforms rely on relational databases (e.g., PostGIS), which become highly inefficient at large data sizes (>10 terabytes), or are based on files (geotiff, ESRI shapefile, etc.) in object store or file systems, whose content is very difficult and slow to search. Consider, for example, a search for the time-series surface reflectance in satellite data files. Doing this today with satellite image repositories is quite involved, and may require searching for the available tiles, and then downloading, resampling, reprojecting, and opening each one to extract pixels within the area of interest (AOI). Alternative approaches, which go beyond file-based indexing, have been proposed, but are still not demonstrated on a large scale.

The second distinguishing characteristic of PAIRS Geoscope is the availability of hierarchical resolution levels to accommodate for different spatial and temporal resolutions, thereby linking the different layers of geospatial-temporal information. For the spatial domain, in the current system, the keys are arranged in z-order, where the length of the key determines the resolution, which ranges from centimeters to kilometers. Extending this concept to the temporal domain, adding more information to the key, optimizing the order of the information in the key, etc., are current research goals.

The third characteristic of PAIRS Geoscope is the alignment of all data to a global grid and the organization of it in linked layers *at the time of ingestion*, contrary to ingestion on demand, which enable superior and scalable performance plus rapid response to queries involving multiple layers. For such ingestion, key design and supercell design must be optimized. In fact, PAIRS Geoscope enables complex analytics during the query. In addition to filtering and aggregating across time and space, user-defined functions can be included in a query that allows for arbitrary mathematical manipulations between layers. An example would be a machine-learnt decision tree, which can be submitted within the query. Finally, PAIRS includes Apache Spark and GPU nodes on the PAIRS Geoscope core cluster to host query results and enable users to interact with these results using a

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 01,2024 at 03:25:00 UTC from IEEE Xplore. Restrictions apply.

containerized Python environment for customized analytics; this also minimizes data movement.

From a user's perspective, PAIRS Geoscope offers four core services. First, on a very basic level, PAIRS Geoscope provides data services for the 4 PB of curated information. On the next level, search or query services can be used to run basic analytics on the PAIRS Geoscope data; for example, filtering different layers of geospatial-information or aggregations. For users who want to develop custom analytics, PAIRS provides analytics platform services such as a containerized, isolated Python environment. Finally, users can upload their own data using the data curation and ingestion services. This allows users to run scalable search and analytics on their data along with the other 4 PB of PAIRS Geoscope data.

Disaster monitoring is a Big Data problem

One area where GIS-based data has recently gained traction is in disaster management. The frequency and spatial extent of natural disaster events are steadily increasing driven by weather and climate change [11]. Earthquakes, heatwaves, flooding, wildfires, tornadoes, landslides, and power outages [12] all have a strong spatial-temporal component. On a global scale, using open-source satellite data like Landsat or Modis across multiple years, forest [13], water [14], and urban areas [15] have been mapped to establish baseline GIS layers for every location on the Earth. Change detection is a powerful tool to quantify the impact of a disaster event, comparing the most current imagery, during or shortly after the event, to a historical snapshot of the same area.

Access to historical imagery and IoT data as they are stored in PAIRS Geoscope can enable a timely analysis to establish the post-disaster event signal deviation from the "normal" baseline. Quick access to historical data trends is important as many events happen with little or no warning, taking people and the authorities by surprise. Anomalies in the GIS data caused by natural disaster events require a quick, on-the-fly filtering of multiple data layers on a computational platform that retrieves only data for the AOI, avoiding time-consuming processing of file stored images. Continental-scale data such as weather models to track hurricane paths can be combined with local data to quickly compare areas that are separated by hundreds of miles but affected by the same event. In the case of earthquakes or landslides, the events happen almost instantaneously, but the results can be influenced by long-term effects like precipitation, soil moisture, and state of vegetation. This type of analysis requires a rapid assessment of the "normal" state in a certain location followed by a quick overlay of new GIS data as they became available after the event. Other events like drought or ocean level rising can develop in a season, requiring a longer baseline to see subtle changes across multiseason or yearly time intervals. In many disaster scenarios, timely access to information about the extent and impact of events is critical [16]. In the following, we will focus on the special case of earthquake and wildfire events as two use cases, where the interplay between sensing, modeling, and early warning can significantly reduce the "time to response" of government agencies.

Disaster monitoring and emergency response management

During the first few hours, after a disaster event, there is usually a lack of information about the extent, impact, and severity of the damage. While the event may be broadcast by news outlets, there is a lack of information in the form of imagery or video. For the few hours after the disaster event, an untapped data stream can be local sensor data, like embedded sensors in infrastructures, mobile devices, or social data streams. These local data sources, while they currently exist, are stored in multiple databases and platforms, requiring extensive processing to spatially align and fill in the gaps. For first responders and aid agencies, it is important to quickly combine the new data streams with established GIS baseline layers to track changes. A framework like PAIRS Geoscope enables the spatial and temporal linking of the IoT data with GIS baseline layers to quickly assess the impact of an earthquake or other disaster.

The importance of GIS data can be ranked according to how quickly the information becomes available after an event.

During and shortly after an earthquake, the only available data are from ground-based sensors like the seismic sensor network operated by the U.S. Geological Survey (USGS). The sensors pick up earth vibrations and use signal triangulation to localize and to quantify the magnitude of the event. In general, data from more than three stations are necessary to localize the epicenter, consequently requiring the operation of a dense sensor network that is evenly spaced across the globe. The information generated by the sensor network will be available in seconds or minutes after the event. These dedicated earthquake sensors can be augmented by less precise but more cost-effective sensors on public mobile platforms, where a dense deployment enables sensing close to the epicenter [17, 18].

Social media is emerging as a quick information platform where the impact of a disaster is documented by crowdsourcing. While the accuracy of the information may be low, the spatial coverage and density can be high as many people might be affected [19].

Remote sensing and drone imagery can highly increase the spatial and temporal coverage of a disaster event. These data can be quickly integrated into geophysical models to assess the extent and severity of the damage [20] based on earthquake wave propagation, vulnerability of buildings

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 01,2024 at 03:25:00 UTC from IEEE Xplore. Restrictions apply.

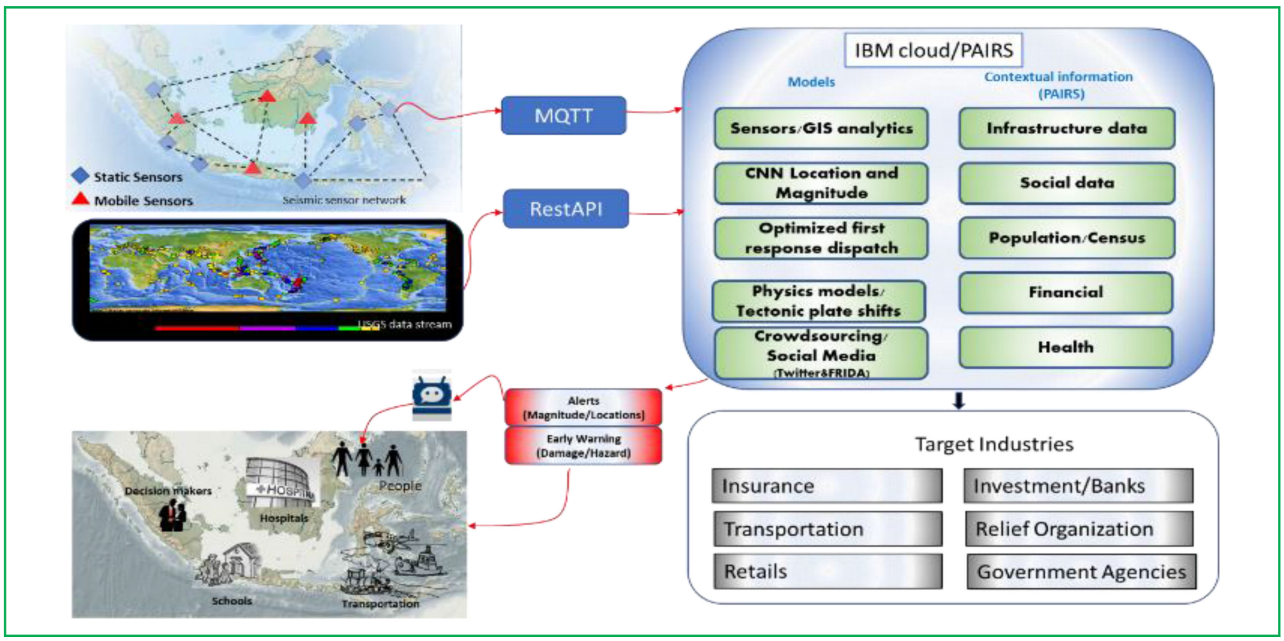

Figure 2

Early warning and information distribution system using multiscale GIS data processing and information dissemination.

and infrastructure, and local conditions like terrain and construction patterns. Models can also be statistical, based on good knowledge of what is on the ground and how it is impacted by vibrations.

Field surveys can be carried out by rescue teams deployed by government agencies to visually inspect buildings and infrastructure and build up knowledge based on special assessments, drone, and airplane imagery [21]. These information sources require extensive processing to convert professional assessments into action.

Long-term modeling of historical events and predictive models that are general in nature can suggest trends and the impact of similar events.

Given the variety of information created during and after a disaster event, the GIS platform needs to quickly ingest all the possible data sources, align as well as validate the data sources, and carry out extensive analysis to assess the impact. Many of the above data sources are input in damage models that can quickly quantify the extent of damage across an area. These models can be physics, statistical, or data-driven. Recently, it has been demonstrated that machine-learned models have similar accuracy to GIS-based models [22]. Besides sensor data, semantic information from social media that may have a roughly localized GIS component may be an input to models [23, 24]. The PAIRS Geoscope's advantage is the ability to quickly search across multiple data layers and integrate the extracted information into a decision tool. The disaster-related data can be quickly combined with census information related to the number of people, houses, and critical infrastructure that may be affected within the area. This information can direct first-response teams and help in the rescue and extraction of survivors. Access to the most recent and accurate GIS data [25, 26] can enable this assessment to be partially automated such that data processing is triggered by the severity of a disaster event like an earthquake. Then, all these information and model results are distributed to stakeholders (see **Figure 2**). Access to timely data like shelter location, potable water sources and availability, and first aid stations could be reduced to minutes instead of hours.

The data flow architecture of an early warning system based on PAIRS Geoscope is shown in Figure 2. The data distribution includes target industries and stakeholders like public, first response teams, public health institutions, transportation, insurance, retail, and government who may need to change their business practice depending on the collected information. Information from either IoT-based sensor networks or remote sensing data are continuously combined into early warning and data distribution systems, while at the same time, they can enable on-the-fly modeling to extract potential damage and provide alerts.

Wildfire monitoring

There are numerous ways to model and monitor events in the context of disaster management. Here, we focus on wildfire detection as an example of combining

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 01,2024 at 03:25:00 UTC from IEEE Xplore. Restrictions apply.

heterogeneous data sources. Considering a fire life cycle, one might think about physical and statistical models of ignition, the availability and distribution of burnable material, spread of the fire and smoke and their effects, management of ongoing fires, and finally, mitigation. While there is ongoing research into all these directions and more, one might argue that the central problem to all of them is a good understanding of the ignition and spread of a wildfire: Any effort to manage an ongoing fire requires reliable estimates of its future path. In what follows, we will focus on models of wildfire spread. We are especially interested in the extent to which research efforts have been or will be affected by the emergence of Big Data and the machine- and deep-learning techniques they facilitate.

Rothermel's work [27] and its subsequent refinement by Albini [28] constitute an early yet hugely impactful example of a wildfire model. Note that Rothermel focuses on surface fires since "mechanisms of heat transfer in a crown fire are considerably different than those for a ground fire." In a nutshell, a model such as Rothermel's gives a number of semi-empirical equations that reflect the fire's spread in terms of various input parameters. These input parameters are essentially concerned with the distribution and properties of fuels as well as environmental parameters such as wind and moisture. Due to the vast variety of fuels occurring in nature as well as the different states they can be in, it is generally necessary to encapsulate the fuels in a fuel model.

It is striking to note that Rothermel's work still constitutes the basis of a large number of modern wildfire models that are under active development. Indeed, Andrews writes in 2018 [29] that "the Rothermel surface fire spread model is the most commonly used in U.S. fire management systems, with a significant use outside the United States." For a wider outlook regarding the model landscape, consider Sullivan's reviews of model developments between 1990 and 2007 [30–32].

As exemplified by its dependence on spatial and temporal input distributions of fuel and environmental parameters, wildfire modeling is highly dependent on geospatial-temporal data. Indeed, a problem of ongoing active research is the characterization of vegetation and fuel distributions [33, 34]. Given progress in large-scale geospatio-temporal data processing as well as machine learning, one immediately wonders to what extent one can leverage purely data-driven methods to approach the issue of wildfire modeling, or that of fuel distributions, for that matter. After all, by considering the temporal evolution of the wildfire distribution given both the fuel distribution as well as external variables (and possibly statistical forecasts for the future evolution of these external variables), wildfire propagation can be rephrased as a supervised learning problem and thus be amenable to deep learning. Surprisingly though, any direct application of deep learning to fire propagation seems to be limited to a small number of studies [35]. Very similar approaches have been explored for flood monitoring [36, 37] and earthquake monitoring [24].

Table 1 Satellite imagery extracted spectral indices to characterize vegetation burn after a wildfire.

Index	Name	Formula
Normalized Difference Vegetation Index	NDVI	$\frac{(\mathrm{RED}-\mathrm{NIR})}{(\mathrm{RED}+\mathrm{NIR})}$
Normalized Difference Water Index	NDWI	$\frac{(\mathrm{NIR}-\mathrm{sSWIR})}{(\mathrm{NIR}+\mathrm{sSWIR})}$
Normalized Burned Ratio	NBR	$\frac{(\mathrm{NIR}-\mathrm{lSWIR})}{(\mathrm{NIR}+\mathrm{lSWIR})}$
Burned Area Index	BAI	$\frac{1}{(0.1-RED)^2+(0.06-NIR)^2}$
Normalized Burned Ratio Thermal	NBRT	$\frac{(\mathrm{NIR}-\mathrm{lSWIR}(\frac{\mathrm{Thermal}}{1000}))}{(\mathrm{NIR}+\mathrm{lSWIR}(\frac{\mathrm{Thermal}}{1000}))}$

Remote sensing-based wildfire monitoring has been proposed as one scalable way to track global-scale wildfires. The European Union system EFFIS[1] and National Aeronautics and Space Administration system[2] provide operational data for wildfire tracking. Information carried by spectral bands is a good indicator of phenological changes that happened to vegetation after a wildfire event. The most common approach is to use the near-infrared and short-wave infrared wavelength as they are sensitive to the biomass, thus indicating a plant's health. Many of the common satellites like Sentinel 2, Landsat, and Modis have spectral bands in the infrared spectrum. Furthermore, MODIS, the Visible Infrared Imaging Radiometer Suite, and the Geostationary Operational Environmental Satellite have detectors in the thermal band that can detect anomalies due to wildfires. All these data sources provide information at the spatial scale of 500 m or above. Processing Landsat or Sentinel 2 data can improve the spatial resolution down to 10 m and leverage the multiple spectral bands that these satellites carry.

A few of the spectral indices that are studied in the context of vegetation burn are described in **Table 1** [38]. Different indices will capture subtle changes in the vegetation state and the impact and severity of the wildfire. The most commonly used indices are normalized difference vegetation index (NDVI), normalized difference water index (NDWI), and normalized burn ratio (NBR) that are calculated as a combination of spectral bands in visible (RED), near infrared, and short-wave infrared (SWIR). The subtle difference between NDWI and NBR is the wavelength of the short-wave infrared band used for calculations; sSWIR is sensitive around 1.6 μm, while lSWIR is sensitive around 2.2 μm. A common approach for

[1][Online]. Available: (http://effis.jrc.ec.europa.eu/)

[2][Online]. Available: (https://firms.modaps.eosdis.nasa.gov)

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 01,2024 at 03:25:00 UTC from IEEE Xplore. Restrictions apply.

Table 2 Severity of wildfire burned area as classified based on differential Normalized Burned Ratio to classify wildfire impact.

Severity level	dNBR range
Enhanced regrowth, high	-0.5 to -0.251
Enhanced regrowth, low	-0.25 to -0.101
Unburned	-0.1 to 0.099
Low severity	0.1 to 0.269
Moderate-low severity	0.27 to 0.439
Moderate-high severity	0.44 to 0.659
High severity	>0.66

wildfire impact assessment is to calculate the difference of a given vegetation index from after and before the wildfire event. The differential Normalized Burned Ratio (dNBR) can be expressed as follows:

$$dNBR = (NBR)_{after} - (NBR)_{before}. \quad (1)$$

The dNBR values range from –1 to 1, where a value of –1 represents an area that became richer in vegetation, while a value around 1 indicates that vegetation vanished from that area. Taking a snapshot of NBR before and after a wildfire event, dNBR is classified in six groups ranging from vegetation growth to severe vegetation burning. The change in dNBR can be used to delineate the total burned area for regions where ground validation may be missing. In **Table 2**, the severity is ranked based on a USGS proposed range [39]: NBR indices can be calculated at 20 m resolution for Sentinel 2, 60 m for Landsat, and 250 m for MODIS data. Extensive studies in Australia demonstrated that different spectral indices (NDVI, NDWI, NBR) may be more indicative for different forest species [38]. Modis provides a monthly operationally calculated burned area (MCD64A1), and recently Landsat provides a similar dataset [40].

One example of wildfire normalized burned ratio detection in PAIRS is shown in **Figure 3(a)** for the region around Santa Paula, CA, USA, for the Thomas wildfire. The Thomas wildfire was initially detected on December 6, 2017, and contained on January 12, 2018. The dNBR is calculated from Sentinel 2 acquired at December 5, 2017, and January 12, 2018, and the red pixels show the area with high-severity burning. In addition, **Figure 3(b)** illustrates a snapshot from PAIRS with a time series of NDVI values extracted for an area having highly severe burning. The time series indicates that vegetation vanished from the area and is slowly regrowing after the wildfire event. Nevertheless, the long-term time series demonstrates that vegetation values are well below historical values and the vegetation is regrowing slowly (upward slope of time series after the event).

The combination of dNBR and NDVI is a good proxy to validate the extent of a wildfire and quantify the severity of the burned regions. Information extracted from multiple spectral indices combined with weather data can provide indications where replanting will be the most effective, and when the vegetation is fully restored.

Earthquake monitoring and modeling

The global seismic sensor network can estimate in a time frame of minutes the epicenter of an earthquake using

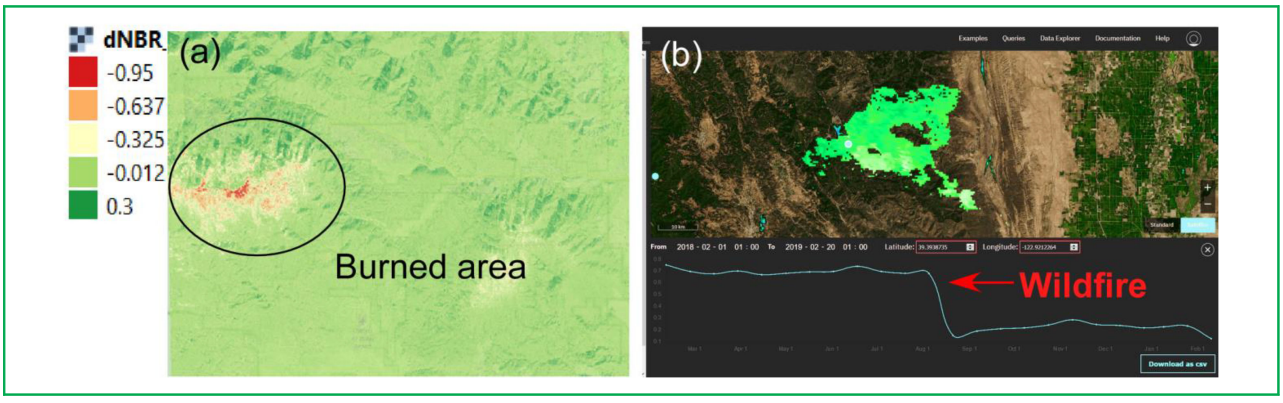

Figure 3

(a) Normalized burned ratio calculation for the Thomas Wildfire around Santa Paula, CA, with red/yellow area showing severe burning. (b) Time trace of the Normalized Difference Vegetation Index for the same area showing an abrupt decrease after the event.

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 01,2024 at 03:25:00 UTC from IEEE Xplore. Restrictions apply.

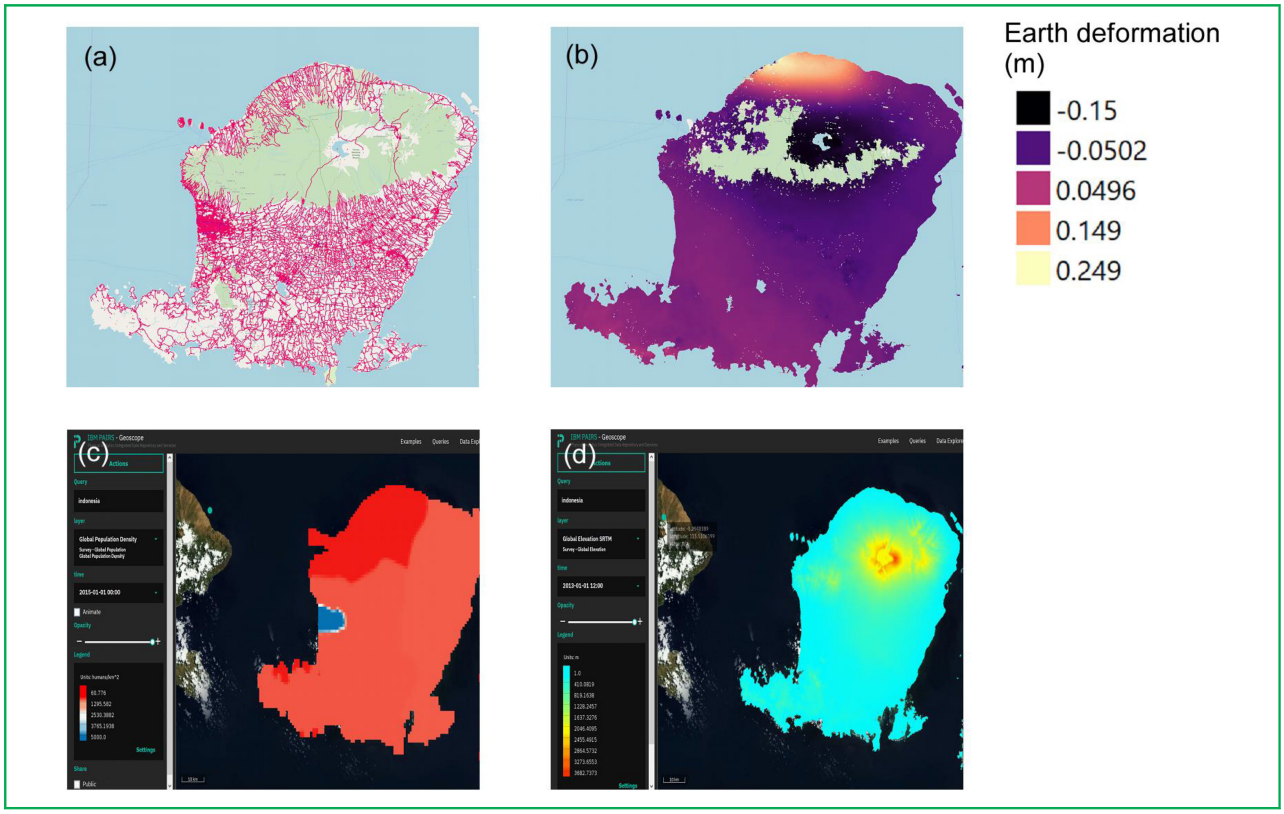

Figure 4

(a) Lombok Island, Indonesia, with the road and bridge network overlaid on a map. (b) Earth deformation after the earthquake on August 5, 2018. (c) PAIRS Geoscope snapshot of population density on the island. (d) PAIRS Geoscope snapshot of elevation.

triangulation from a sensor network. One result of a quake is a deformation of the earth surface resulting in significant changes in terrain and destruction of urban areas and infrastructure. The large-scale damage can be quantified from remote imagery. We present such a scenario based on a 6.8-magnitude earthquake that happened on August 5, 2018, at 11:46 a.m. in Lombok, Indonesia. The epicenter of the Earthquake was 10 km below ground, and the tremble was detected by the USGS seismic sensor network. Synthetic aperture radar (SAR) signals acquired by the European Space Agency were analyzed to assess the impact of Earth surface displacements that commonly occur after an Earthquake. The deformation tends to be a few centimeters in vertical changes that can lead to cracks in roads, buildings, dams, and the Earth's surface. The SAR imagery was acquired on August 5, 2018, hours after the earthquake, and it was compared to a previously acquired snapshot covering the same area. The two images were used to create an interferometry image that quantifies the deformation of the earth from the change in the reflected phase signal. The earth deformation is expressed in centimeters [41]. Preprocessing of the Sentinel 1 data was undertaken using the ESA-SNAP toolbox. The deformation maps show the change in topography height of the Lombok island referenced to the baseline. The largest change is close to the epicenter, and it decreases with distance. The sensitivity of the interferogram image is 2 cm in the vertical direction and 20 m in the horizontal. The images are corrected for terrain-induced artifacts and filtered to minimize noise. The practice to generate interferometry is an established technique for earthquake analysis. The novelty is the fusion of different GIS layers including earth displacement with the distribution of roads, bridges, and population density. From a first aid perspective, reaching the largest number of people affected by an earthquake in the shortest time and knowledge of the available roads and functional bridges can be critical to life-saving efforts. In this context, the classification of road segments with the smallest tilt or deformation can be a proxy for road integrity. Combining multiple data layers as shown in **Figure 4** provides a quick insight into all relevant data that can drive the decision support for providing relief and first aid to the most affected areas. Allocation of resources to areas with higher population densities and the largest

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 01,2024 at 03:25:00 UTC from IEEE Xplore. Restrictions apply.

damage can be the most effective way to mitigate the civic disaster that follows. Analytical tools can use these GIS layers to classify road segments that may be affected by the tilt and shift in elevation from the earthquake. The data can then be enhanced by IoT sensor data that locally quantifies the disturbances. All these analyses require multilayer GIS searching capabilities to quickly scan all relevant layers to discover changes on the ground.

Evaluation of PAIRS geoscope's performance

Disaster prediction and response require readily available data analytics and the ability to run a disaster response model almost on demand, as many disaster events occur without prior indication. Therefore, it is crucial that requested data can be retrieved as quickly as possible in case of such an event for subsequent visualization and analytics. Since geospatial-temporal data is inherently big, retrieval time on such databases is nonnegligible. A concise performance evaluation is presented, comparing PAIRS Geoscope's speed to retrieve data compared with a similar geospatial-temporal data platform that stores the data in a file format and retrieves individual tiles and loads them in computer memory for further processing.

The data *retrieval time* is used as an evaluation metric, where the *retrieval time* is defined as the time difference between submitting a data request to the platform and receiving the actual data back. This time difference includes initialization, queueing, and processing time. Naturally, the retrieval time will depend on the size and complexity of the user's requested data. Any request is formalized by a *query* that defines the specific dataset and layer of interest, the spatial extent, and the temporal range. Moreover, functions to aggregate data across time and space can be specified (e.g., provide the mean value for a time series or for a spatial query or retrieve raster data where the medium value across a time interval is larger than a certain threshold). These query characteristics provide a proxy for the data requests' complexity. For this evaluation, a geospatial-temporal database is more performant if it requires consistently less retrieval time across most of the proposed queries.

A complete enumeration of all possible data queries would serve as an appropriate basis for evaluating a data platform's performance. Unfortunately, this is impossible due to the versatility of query specifications and the fact that only a minority of them would be of practical interest for disaster management. Therefore, for these performance metrics, only a few queries are selected that reflect common yet diverse geospatial-temporal data application scenarios while testing crucial capabilities of the platform. The query selection is guided by theoretical and practical considerations to model spatial characteristics like earthquake-generated damage or to calculate the extent of burned area after a wildfire event.

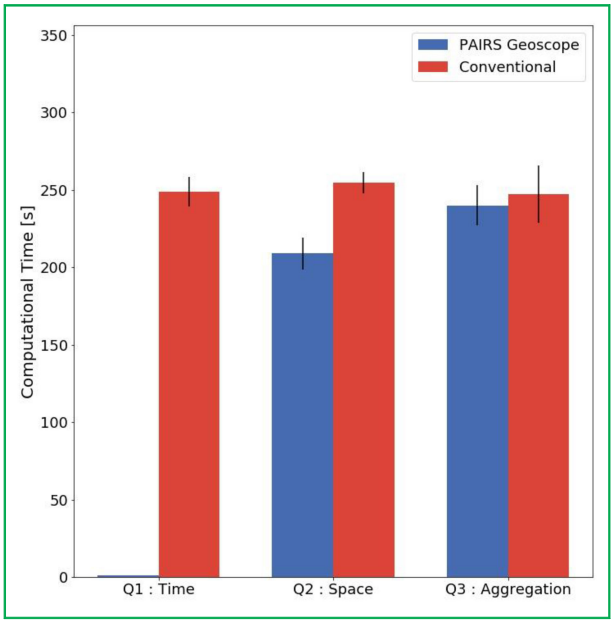

Figure 5

Performance of PAIRS Geoscope compared to a conventional Geospatial Information System for three queries (Q1-time series, Q2-Spatial query, and Q3-Spatial Aggregation on raster data).

In theory, three types of queries are representative to evaluate a platform's overall performance: 1) a query that retrieves a time series for a given set of point coordinates; 2) one that extracts raster information across a given spatial area for a fixed point in time; 3) a query that aggregates data across a time interval for a given spatial area. The former two queries assess a GIS platform's ability to retrieve purely temporal or spatial data, while the third query also measures the platform's computational efficiency to perform statistical computations on the queried data. For each query, data are retrieved according to the database's native resolution and coordinate projection system to exclude computational effort due to sampling or reprojection. The average location and dispersion parameters of the query retrieval times are expressed as median and interquartile range to make the evaluation more robust with respect to potential outliers.

In practice, the three queries (**Figure 5**) are designed to adhere to typical disaster management scenarios presented in previous paragraphs. The specific queries may seem quite general; however, the evaluation results apply to a wide class of similar queries as subsequent empirical investigation reveals. First, the temporal query is specified for wildfire monitoring purposes. More precisely, it retrieves time-series data of NDVI for 25 distinct point coordinates across the Amazon rainforest for one year. This query serves the purpose to track the long-term damage of wildfires on vegetation [42]. Second, the spatial query relates to

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 01,2024 at 03:25:00 UTC from IEEE Xplore. Restrictions apply.

earthquake emergency response, as it provides a single high-resolution image of the island of Lombok. This image, in turn, can be integrated into geophysical models or be combined with other data layers such as population density or infrastructure as shown in Figure 4 to assess disaster response [43]. Finally, a query is submitted that provides the maximum difference of near-infrared images of an area across one year. (Near-) infrared remote sensing is not only relevant for vegetation management, as it is a key ingredient in NDVI, but has been discussed in the context of other natural disasters such as volcano eruptions and landslides [44–46]. For the analysis, the Sentinel 2 dataset has been chosen as it offers high-spatial-resolution imagery in 13 bands including the ones proposed for the query [47].

The performance evaluation, across the three queries, indicates PAIRS Geoscope's advantage to spatially and temporally align all data layers. It performs better across all query types as its median retrieval times are consistently shorter than that of the conventional geospatial-temporal database. The difference is statistically significant for time and space queries, respectively, as indicated by the nonoverlapping error bars. The temporal query on PAIRS Geoscope is processed by a factor of 100 faster than that on the competing platform. This advantage is of major importance as time-series data is ubiquitous for meteorological and climatological geostatistics [48–50].

The conventional platform, on the other hand, requires comparable retrieving time across all queries despite the difference in the requested data types and sizes. Further empirical investigations reveal that these performance results are robust with respect to the choice of dataset and layer. Moreover, the relative performance of the platforms remains unchanged when increasing spatial extent or the number of timestamps, i.e., the effective data size to be processed.

Conclusion

With the readily available GIS data acquired by government agencies, commercial companies, and citizens, there is a need for new data platforms that can easily process different data formats and enable real-time processing and modeling. The variety of data formats ranging from images, text, and point clouds require a comprehensive data management and processing platform. The description and performance characteristics of such a platform called PAIRS is presented, which spatially and temporally aligns all data sources in a Big Data framework. While local measurement based on special sensors in the context of IoT provides the fastest detection of changes on the ground (detecting earthquakes, assessing damage, disseminating the information) and the scalability of this approach is under investigation, remote sensing has the virtue of quickly surveying large areas and observing ground-based changes. The availability of curated data sets that are continuously cross-validated with existing data sources enables quick modeling and data retrieval of all relevant data layers in a disaster scenario. Such large-scale surveys can enable optimization on the ground for first responders and aid agencies to assess damage and quickly allocate resources to regions based on large-scale damage assessment.

Acknowledgment

This work was supported by IBM.

References

1. J. Snow, *On the Mode of Communication of Cholera*. London, U.K.: Churchill, 1855.
2. R. F. Tomlinson, *Thinking About GIS: Geographic Information System Planning for Managers*, vol. 1. Redlands, CA, USA: ESRI, Inc., 2007.
3. P. Soille, A. Burger, D. De Marchi, et al., "A versatile data-intensive computing platform for information retrieval from big geospatial data," *Future Gener. Comput. Syst.*, vol. 81, pp. 30–40, 2018.
4. L. Azaz, "The use of geographic information systems (GIS) in business," in *Proc. Int. Conf. Humanities, Geography, Econ.*, 2011, pp. 299–303.
5. D. Preoţiuc-Pietro, J. Cranshaw, and T. Yano, "Exploring venue-based city-to-city similarity measures," in *Proc. 2nd ACM SIGKDD Int. Workshop Urban Comput.*, 2013, pp. 1–4.
6. L. J. Klein, F. J. Marianno, C. M. Albrecht, et al., "PAIRS: A scalable geo-spatial data analytics platform," in *Proc. IEEE Int. Conf. Big Data*, 2015, pp. 1290–1298.
7. S. Lu, X. Shao, M. Freitag, et al., "IBM PAIRS curated Big Data service for accelerated geospatial data analytics and discovery," in *Proc. IEEE Int. Conf. Big Data*, 2016, pp. 2672–2675.
8. F. Warmerdam, "The geospatial data abstraction library," in *Open Source Approaches in Spatial Data Handling*. New York, NY, USA: Springer, 2008, pp. 87–104.
9. J. N. Hughes, A. Annex, C. N. Eichelberger, et al., "Geomesa: A distributed architecture for spatio-temporal fusion," *Proc. SPIE*, vol. 9473, 2015, Art. no. 94730F.
10. M. A. Whitby, R. Fecher, and C. Bennight, "Geowave: Utilizing distributed key-value stores for multidimensional data," in *Proc. Int. Symp. Spatial Temporal Databases*, 2017, pp. 105–122.
11. J. Rundle, *Reduction and Predictability of Natural Disasters*. Abingdon, U.K.: Routledge, 2018.
12. M. Pelling, *Natural Disaster and Development in a Globalizing World*. Abingdon, U.K.: Routledge, 2003.
13. M. C. Hansen, P. V. Potapov, R. Moore, et al., "High-resolution global maps of 21st-century forest cover change," *Science*, vol. 342, pp. 850–853, 2013.
14. J.-F. Pekel, A. Cottam, N. Gorelick, et al., "High-resolution mapping of global surface water and its long-term changes," *Nature*, vol. 540, 2016, Art. no. 418.
15. Y. Zeng, W. Huang, M. Liu, et al., "Fusion of satellite images in urban area: Assessing the quality of resulting images," in *Proc. 18th Int. Conf. Geoinformat.*, 2010, pp. 1–4.
16. E. C. De Perez, F. Monasso, M. Van Aalst, et al., "Science to prevent disasters," *Nature Geosci.*, vol. 7, 2014, Art. no. 78.
17. E. S. Cochran, J. F. Lawrence, C. Christensen, et al., "The quake-catcher network: Citizen science expanding seismic horizons," *Seismological Res. Lett.*, vol. 80, pp. 26–30, 2009.
18. M. D. Kohler, T. H. Heaton, and M.-H. Cheng, "The community seismic network and quake-catcher network: Enabling structural health monitoring through instrumentation by community participants," *Proc. SPIE*, vol. 8692, 2013, Art. no. 86923X.
19. T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: Real-time event detection by social sensors," in *Proc. 19th Int. Conf. World Wide Web*, 2010, pp. 851–860.
20. R. Spence, E. So, G. Ameri, et al., "Earthquake disaster scenario prediction and loss modelling for urban areas," Coll. Higher Stud. Pavia, Pavia, Italy, LESSLOSS Rep. 2007/07, 2007.

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 01,2024 at 03:25:00 UTC from IEEE Xplore. Restrictions apply.

21. F. Yamazaki and W. Liu, "Remote sensing technologies for post-earthquake damage assessment: A case study on the 2016 Kumamoto earthquake," in *Proc. 6th Asia Conf. Earthquake Eng.*, 2016, pp. 193–210.
22. F. Corbi, L. Sandri, J. Bedford, et al., "Machine learning can predict the timing and size of analog earthquakes," *Geophys. Res. Lett.*, vol. 46, pp. 1303–1311, 2019.
23. T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural network for earthquake detection and location," *Sci. Adv.*, vol. 4, 2018, Art. no. e1700578.
24. P. M. DeVries, F. Viégas, M. Wattenberg, et al., "Deep learning of aftershock patterns following large earthquakes," *Nature*, vol. 560, 2018, Art. no. 632.
25. S. Nevo, V. Anisimov, G. Elidan, et al., "ML for flood forecasting at scale," in *Proc. Thirty-Second Ann. Conf. Neural Inform. Process. Syst. (NIPS)*, Montréal, QC, Canada, Dec. 2018, [Online]. Available: https://arxiv.org/pdf/1901.09583.pdf , Accessed on: 10 Feb. 2020.
26. Y. Gigi, G. Elidan, A. Hassidim, et al., "Towards global remote discharge estimation: Using the few to estimate the many," 2019, *arXiv:1901.00786*.
27. R. C. Rothermel, "A mathematical model for predicting fire spread in wildland fuels," U.S. Dept. Agriculture, Intermountain Forest Range Exp. Station, Ogden, UT, USA, Res. Pap. INT-115, 1972.
28. F. A. Albini, "Estimating wildfire behavior and effects," U.S. Dept. Agriculture, Forest Service, Intermountain Forest Range Exp. Station, Ogden, UT, USA, Gen. Tech. Rep. INT-GTR-30, 1976.
29. P. L. Andrews, "The Rothermel surface fire spread model and associated developments: A comprehensive explanation," U.S. Dept. Agriculture, Forest Service, Rocky Mountain Res. Station, Fort Collins, CO, USA, Gen. Tech. Rep. RMRS-GTR-371, 2018.
30. A. L. Sullivan, "Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models," *Int. J. Wildland Fire*, vol. 18, pp. 349–368, 2009.
31. A. L. Sullivan, "Wildland surface fire spread modelling, 1990–2007. 2: Empirical and quasi-empirical models," *Int. J. Wildland Fire*, vol. 18, pp. 369–386, 2009.
32. A. L. Sullivan, "Wildland surface fire spread modelling, 1990–2007. 3: Simulation and mathematical analogue models," *Int. J. Wildland Fire*, vol. 18, pp. 387–403, 2009.
33. M. G. Rollins, "LANDFIRE: A nationally consistent vegetation, wildland fire, and fuel assessment," *Int. J. Wildland Fire*, vol. 18, pp. 235–249, 2009.
34. R. Nair and S. Gupta, "Wildfire: Approximate synchronization of parameters in distributed deep learning," *IBM J. Res. Dev.*, vol. 61, no. 4/5, pp. 7-1–7-9, Jul./Sep. 2017.
35. P. Zohouri Haghian, *Deep Representation Learning and Prediction for Forest Wildfires*. Waterloo, ON, Canada: Univ. Waterloo, 2019.
36. H. Shafizadeh-Moghadam, R. Valavi, H. Shahabi, et al., "Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping," *J. Environ. Manage.*, vol. 217, pp. 1–11, 2018.
37. A. Mosavi, P. Ozturk, and K.-W. Chau, "Flood prediction using machine learning models: Literature review," *Water*, vol. 10, 2018, Art. no. 1536.
38. B. Tran, M. Tanase, L. Bennett, et al., "Evaluation of spectral indices for assessing fire severity in Australian temperate forests," *Remote Sens.*, vol. 10, 2018, Art. no. 1680.
39. C. H. Key and N. C. Benson, "Landscape assessment (LA)," in FIREMON: Fire Effects Monitoring and Inventory System, U.S. Dept. Agriculture, Forest Service, Rocky Mountain Res. Station, Fort Collins, CO, USA, Gen. Tech. Rep. RMRS-GTR-164-CD, 2006.
40. T. J. Hawbaker, M. K. Vanderhoof, Y.-J. Beal, et al., "Mapping burned areas using dense time-series of Landsat data," *Remote Sens. Environ.*, vol. 198, pp. 504–522, 2017.
41. S. Plank, "Rapid damage assessment by means of multi-temporal SAR—A comprehensive review and outlook to Sentinel-1," *Remote Sens.*, vol. 6, pp. 4870–4906, 2014.
42. M. A. Cochrane, "Fire science for rainforests," *Nature*, vol. 421, 2003, Art. no. 913.
43. N. Kerle, "Satellite-based damage mapping following the 2006 Indonesia earthquake—How accurate was it?," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 12, pp. 466–476, 2010.
44. C. Wei, Y. Zhang, X. Guo, et al., "Thermal infrared anomalies of several strong earthquakes," *Sci. World J.*, vol. 2013, 2013, Art. no. 208407.
45. R. Prakash and H. Srivastava, "Thermal anomalies in relation to earthquakes in India and its neighbourhood," *Current Sci.*, vol. 108, pp. 2071–2082, 2015.
46. D. M. Tralli, R. G. Blom, V. Zlotnicki, et al., "Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards," *ISPRS J. Photogrammetry Remote Sens.*, vol. 59, pp. 185–198, 2005.
47. M. Drusch, U. Del Bello, S. Carlier, et al., "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012.
48. K. Krivoruchko, *Spatial Statistical Data Analysis for GIS Users*. Redlands, CA, USA: Esri Press, 2011.
49. C. W. Landsea, "Meteorology: Hurricanes and global warming," *Nature*, vol. 438, 2005, Art. no. E11.
50. C. Funk, A. Verdin, J. Michaelsen, et al., "A global satellite assisted precipitation climatology," *Earth Syst. Sci. Data Discuss.*, vol. 8, pp. 275–287, 2015.

Received December 10, 2018; accepted for publication October 10, 2019

Conrad M. Albrecht *IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (cmalbrec@us.ibm.com).* Dr. Albrecht received an undergraduate degree in physics from Technical University Dresden, Dresden, Germany, in 2007 and a Ph.D. degree in physics with an extra certification in computer science from Heidelberg University, Heidelberg, Germany, in 2014. He is currently a Research Scientist with the IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. Spanning the fields of physics, mathematics and computer science, among others, he was a Visiting Scientist with CERN, Switzerland, in 2010, and the Max Planck Institute for the Physics of Complex Systems, Dresden, in 2007. His research interests include interconnecting physical models and numerical analysis, employing Big Data technologies. As part of the Data Intensive Physical Analytics team in IBM Research, he is working on Big Data processing for geospatial data analysis and machine-learning driven remote sensing. He is a Member of the IEEE and Co-Organizer of workshops at the IEEE BigData conference. As an open-source advocate, he maintains the Python software development kit for IBM PAIRS Geoscope: https://github.com/IBM/ibmpairs.

Bruce Elmegreen *IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (bge@us.ibm.com).* Dr. Elmegreen received an undergraduate degree from the University of Wisconsin in Madison, Madison, WI, USA, in 1971 and a Ph.D. degree in astrophysics from Princeton University, Princeton, NJ, USA, in 1975. He is currently in the Physical Analytics and IoT Department of the Research Division of IBM, Yorktown Heights, NY, USA. He was a Junior Fellow with Harvard University. In astrophysics, he studies star formation and galactic structure, having written more than 360 scientific articles. He was the Chair of the Publications Board of the American Astronomical Society (AAS, 1998–2001) and is currently a member of the Publications Board of the Astronomical Society of the Pacific. He was the recipient of the Dannie Heineman Prize of the AAS and the American Physical Society in 2001, and an IBM Research Outstanding Accomplishment Award for Research in Star Formation in 2016. From 2015–2018, he was the President of Division H on Interstellar Matter and Local Universe and Chair of the Resolutions Board for the International Astronomical Union (IAU), and he was the National Representative of the AAS to the IAU. In the field of nanotechnology, he has designed and modeled magnetic materials and devices, phase change material structures, and piezoelectronic devices and circuits, obtaining 23 patents. He has also modeled traffic flow in cities and studied long-term weather patterns. He was a member of the U.S. National Science Foundation Math and Physical Sciences Advisory Committee from 2012 to 2016 and became a Fellow of the American Association for the Advancement of Science in 2013.

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 01,2024 at 03:25:00 UTC from IEEE Xplore. Restrictions apply.

Oki Gunawan *IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (ogunawa@us.ibm.com).* Dr. Gunawan received a B.Eng. degree and an M.Eng. degree from Nanyang Technological University, Singapore, in 1998 and 2000, respectively, and an M.A. degree and a Ph.D. degree from Princeton University, Princeton, NJ, USA, in 2007, all in electrical engineering. He is currently a Research Staff Member with IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. His research interests include semiconductor technology and physics such as nanowire transistor, photovoltaics, electrodynamics, and novel sensors for Internet of Things (IoT). At IBM, he is responsible in managing device characterization efforts in the IBM CZTS (CuZnSnSeS) solar cell project that achieved world-record efficiency in 2013. He has authored and coauthored more than 50 publications and holds 51 U.S. and international patents. He was the recipient of the IBM Master Inventor Award in 2019, the Achmad Bakrie Award in 2013 (Indonesia, Young scientist category), the IBM Invention Achievement Award (9th plateau), the IBM Research Division Award in 2010, and the IBM New Hire Patent Award in 2009.

Hendrik F. Hamann *IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (hendrikh@us.ibm.com).* Dr. Hamann received a Ph.D. degree from the University of Göttingen, Göttingen, Germany, in 1995. He is currently a Senior Manager and Distinguished Research Staff Member with the IBM T. J. Watson Research Center, Yorktown Heights, NY, USA, which he joined in 1999. His research interests include sensor networks, sensor-based physical modeling, machine learning, artificial intelligence, Big Data technologies, and geospatial analytics. He has authored and coauthored more than 140 peer-reviewed scientific papers and holds more than 130 patents. He is an IBM Master Inventor and a member to the IBM Academy of Technology, and has served on governmental committees such as the National Academy of Sciences and the National Science Foundation, and as an industrial advisor to universities. He was the recipient of several awards including the 2016 AIP Prize for Industrial Applications of Physics. He is a member of the American Physical Society (APS), the Optical Society of America (OSA), The Institute of Electrical and Electronics Engineers (IEEE), and the NY Academy of Sciences.

Levente J. Klein *IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (kleinl@us.ibm.com).* Dr. Klein received a Ph.D. degree in physics from the University of Utah, Salt Lake City, UT, USA, in 2002. He is currently a Research Staff Member in the Internet of Things and Industry Solution with the IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. His work at IBM spans multiple research topics such as material science, nano-optics, and wireless sensing solutions with strong focus to apply research technologies to industrial problems. Since joining IBM Research in 2006, he developed technologies to enable energy-efficient cooling in data centers, monitoring fugitive methane gas in oil and gas industry, and application of wireless sensing solution in agriculture and healthcare. His current research interests include environmental monitoring of greenhouse gases, application of wireless sensing in outdoor environment, and physics-based modeling and analytics. He was the recipient of the three IBM Outstanding Technical Achievement Awards and multiple IBM Research Division Awards.

Siyuan Lu *IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (lus@us.ibm.com).* Dr. Lu received a Ph.D. degree in physics from the University of Southern California (USC), Los Angeles, CA, USA, in 2006. He is currently a Manager and Principle Research Staff Member with the IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. Prior to joining IBM in 2012, he was an Assistant Research Professor jointly appointed in the Department of Physics and the Department of Ophthalmology at USC. His research interests include architectures of geospatial data services and the union of physics and data-driven approaches for modeling complex systems. He currently leads the Data Intensive Physical Analytics research group that focuses on the development of IBM PAIRS Geoscope Big Data services and its applications in renewable energy forecasting, climate forecasting, and remote-sensing-based land surveying, and agricultural and environmental monitoring. He has coauthored more than 50 peer-reviewed articles and has served as a member on multiple governmental committees.

Fernando J. Marianno *IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (fjmarian@us.ibm.com).* Mr. Marianno received a B.Sc. degree in computer science from the Methodist University of Piracicaba, Piracicaba, Brazil, in 2001. He is currently a Senior Software Engineer with the Science and Technology Department, IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. He has authored or coauthored nine patents and ten technical papers. He was the recipient of an IBM Outstanding Technical Achievement Award for developing measurement and management technology in 2013 and another Outstanding Technical Award in 2015 for his contributions to the corrosion monitoring on pSeries and zSeries platforms. He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and the American Geophysical Union (AGU).

Carlo Siebenschuh *IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (Carlo.Siebenschuh@ibm.com).* Mr. Siebenscuh received two master's degrees in pure mathematics and industrial engineering, respectively, from the Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2017. He is currently a Software Engineer with the IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. His research interests include intersecting mathematical statistics and machine learning to develop and apply new learning algorithms to large-scale datasets. In the past, he worked on high-dimensional data arising in finance and health. In 2018, he joined the physical analytics group and has since focused working on geospatial-temporal data emerging in remote sensing.

Johannes Schmude *IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (johannes.schmude@ibm.com).* Mr. Schmude received a Ph.D. degree from the University of Swansea, Swansea, U.K., in 2010. He is currently a Research Staff Member with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA. He joined the Kavli Institute for the Physics and Mathematics of the Universe, University of Tokyo, as a JSPS fellow. Following further postdoctoral appointments in RIKEN's mathematical physics group and in the high energy physics group at the University of Oviedo, he joined IBM in 2017. His early publications focused on the intersection of high energy theory and differential geometry. His current work concerns problems in geospatial Big Data, machine learning, and artificial intelligence.

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 01,2024 at 03:25:00 UTC from IEEE Xplore. Restrictions apply.