Geospatial sensor web: A cyber-physical infrastructure for geoscience research and application

Xiang Zhang\textsuperscript{a}, Nengcheng Chen\textsuperscript{a,b,*}, Zeqiang Chen\textsuperscript{a}, Lixin Wu\textsuperscript{c}, Xia Li\textsuperscript{d}, Liangpei Zhang\textsuperscript{a,b}, Liping Di\textsuperscript{e}, Jianya Gong\textsuperscript{a,b}, Deren Li\textsuperscript{a,b}

\textsuperscript{a} State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan 430079, China
\textsuperscript{b} Collaborative Innovation Center of Geospatial Technology, Wuhan 430079, China
\textsuperscript{c} School of Geosciences and Info-Physics, Central South University, Changsha 410083, China
\textsuperscript{d} School of Geographic Sciences, Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai 2000241, China
\textsuperscript{e} Center for Spatial Information Science and Systems, George Mason University, Fairfax 22030, USA

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Collaborative observation
Focusing service
Geospatial Infrastructure
Integrated management
Scalable processing
Sensor web

\textbf{ABSTRACT}

In the last half-century, geoscience research has advanced due to multidisciplinary technologies, among which Information and Communication Technology (ICT) has played a vital role. However, scientifically organizing these ICTs toward improving geoscience measurements, data processing, and information services has encountered tremendous challenges. This paper reviews a profound revolution in geoscience that has resulted from the Geospatial Sensor Web (GSW), serving as a new cyber-physical spatio-temporal information infrastructure for geoscience on the World Wide Web (WWW). In contrast to previous experiment-based and sensor-based paradigms, the GSW-based paradigm is able to accomplish the following: (1) achieve integrated and sharable management of diverse sensing resources, (2) obtain real-time or near real-time and spatiotemporal continuous data, (3) conduct interoperable and online geoscience data processing and analysis, and (4) provide focusing services with web-based geoscience information and knowledge. As a benefit of the GSW, increasingly more geoscience disciplines are enjoying the value of real-time data, multi-source monitoring, online processing, and intelligence services. This paper reviews the evolution of geoscience research paradigm to demonstrate the scientific background of GSW. Then, we elaborate on four key methods provided by GSW, namely, integrated management, collaborative observation, scalable processing and fusion, and focusing service web capacity. Furthermore, current GSW prototypes and applications for environmental, hydrological, and natural disaster analysis are also reviewed. Moreover, four challenges to the future GSW in geoscience research are identified and analyzed, including integration with the Model Web initiative for sophisticated geo-processing, integration with humans for pervasive sensing, integration with Internet of Things (IoT) to achieve high-quality performance and data mining, and integration with Artificial Intelligence (AI) to provide smart geoservices. We have concluded that GSW has become an indispensable cyber-physical infrastructure, and will play a greater role in geoscience research and application.

\section{1. Introduction}

Geoscience is a widely used term in earth science fields. Geoscientists face a unique challenge in seeking to understand the complexity of the Earth’s physical and biogeochemical systems. In particular, surface environment of the Earth is controlled by complex interactions between the deep Earth, the atmosphere, the hydrosphere, and the biosphere. These interactions occur on timescales ranging from picoseconds, for chemical reactions on mineral surfaces, to the billions of years, during which plate tectonic processes and biological evolution have radically altered the composition of the atmosphere (Princeton University, 2014). Therefore, the key in geoscience research is to develop highly efficient measurement and analytical technology (Allegre and Courtillot, 1999; Bi, 2004; Lautenbacher, 2005; Reid et al., 2010). The most traditional method adopted by geoscientists is to obtain samples/measurements on the spot, which are then taken back to the laboratory for chemical/physical/biological analyses. This experiment-based approach is workable, especially in the early stages. However, it...
is a labor-intensive approach and is not suitable for rapidly changing or large-scale phenomena, such as earth surface disasters.

To promote geoscience research, satellite Remote Sensing (RS) and Environmental Sensor Networks (ESN) are considered two revolutionary research tools (Hart and Martinez, 2006). RS technology provides geoscientists, for the first time, with a spatially continuous measurement of the atmosphere (Kaufman et al., 2002; Chand et al., 2009; Zou et al., 2016), the earth's surface (Bills, 2009; Hu et al., 2014) and even the underground (Mansor et al., 1994). Whereas ESN allows geoscientists to have more automatic and continuous, but spatially discrete samples of the environment they study (Welsh, 2010). With the help of these technologies, earth science research has advanced in efficiency and scope. However, these sensors, no matter satellite or ground platform, are isolated sensing resources without cooperation. Moreover, most RS-based observations are provided as non-real-time data in isolated databases. Therefore, the processes of Earth systems cannot be fully captured and analyzed, let alone online services for geoscience researchers, government, and the public.

In the last two decades, a Geospatial Sensor Web (GSW) has been proposed as a fundamental paradigm shift in geoscience research (Di, 2007; Chen et al., 2014a; Gong et al., 2015). GSW uses a sensor web that retrieves real-time data from the physical world for the first time on a large scale (Butler, 2006; Di, 2007). Furthermore, GSW is defined as cyber-physical spatiotemporal information infrastructure for geoscience research that benefits from heterogeneous web-ready sensing resources, scalable online processing, and focusing web services. Typically, a cyberinfrastructure integrates advanced computer, information, and communication technologies to empower computation-based and data-driven scientific practices and to improve the synthesis and analysis of scientific data in a collaborative and shared fashion (Yang et al., 2010; Wright and Wang, 2011). In this context, GSW is established as a cyberinfrastructure for geoscience research and application. This is a rebirth of old geoscience research that is now powered by state-of-the-art Information and Communication Technology (ICT).

Because of the significant role played by GSW in current and future geoscience domain, there is an urgent need for a detailed review to add clarity to GSW. The present work not only traces the evolutionary course of geoscience research approaches in the past half-century but also promotes the development of GSW in the coming decades. Examining nearly twenty years of continuous exploration, this article provides a comprehensive and critical review of GSW in geoscience research and application. This review, for the first time, includes the reason for the emergence, advantageous methods, distinct features, prototypes, applications, and future directions of GSW. The remainder of this article includes the following sections: the evolutionary process, from an experiment-based to a GSW-based approach, is illustrated in Section 2. Section 3 identifies four key methods for understanding GSW. Section 5 discusses the implementation and applications of GSW. Finally, the remaining challenges and future perspectives are analyzed in Section 6. The article concludes with a summary of GSW in future geoscience research scenarios.

2. Evolution of the geoscience research paradigm

This section reviewed three typical research paradigms in geoscience field, including experiment-based, sensor-based, and GSW-based paradigm. This evolution was illustrated in Fig. 1, and discussed below. It is also emphasized that no single paradigm can fit all due to the complexity of geoscience scenarios at present, so they will coexist for a long time.

2.1. Experiment-based paradigm

In the early stages, i.e., before the 1970s, most geoscience researchers relied on the experiment-based paradigm, such as field sampling/recording in geology, glaciology, hydrology, mineralogy, paleontology, and geochemistry (Garrett, 1969; McBratney et al., 1981). Typically, a geoscientist selects a representative location and then, conducts in situ measurements. Afterwards, these data/samples are taken back to the laboratory for analysis and recording. Finally, the results are displayed on paper documents or man-made maps.

This paradigm has the following features: (1) it has high accuracy and complete experiment recording; (2) it depends heavily on the knowledge of experts and experiment procedures; (3) this approach requires much labor to collect the raw data, process them, extract useful information, and display them; (4) there is a lag time between the study data and current situation, which means little active response or adaptive management can be conducted on spot; and (5) it cannot guarantee that the conducted experiments effectively capture all weather and around-the-clock reality.

In this paradigm, the experiences of geoscientists are critical for a reliable ground experiment. Even today, there are still some geoscience research scenarios that rely on this experiment-based paradigm. This approach is particularly suitable for some specific geoscience studies which usually involve large equipment (Beaudon et al., 2017), or some parameters that are difficult to be observed by using other remote approaches (Carpenter et al., 2011), or the in-situ sample material that is worth taking back (Filella, 2011).

2.2. Sensor-based paradigm

With the development of new technologies in the computer, information, and electronic fields since the 1970s, increasingly more geoscience disciplines have been renovated with sensors, computers, and network techniques, such as atmospheric science, geoinformatics, seismology, hydrology, and geophysics (Hart and Martinez, 2006). These encompass a sensor-based paradigm in geoscience research. Several significant improvements have been observed.

First, in many cases, geoscientists do not need to take samples back to the laboratory; instead, portal sensors and sensor networks conduct measurements automatically and instantly (Hart and Martinez, 2006; Kim et al., 2012; Batt et al., 2013). For example, high-resolution X-ray computed tomography has become an indispensable technique in the field of geoscience, which is a non-destructive characterization technique that allows for 4D monitoring of internal structural changes at resolutions down to a few hundred nanometers (Snudde and Boone, 2013). Other typical geoscience sensors include thermal infrared spectroscopy on feldspars (Hecker et al., 2010), terrestrial laser scanning (Telling et al., 2017), and spectroradiometric sensors (Debret et al., 2011).

Second, satellite remote sensing first provides estimations for large areas, which is the most important evolutionary development in geoscience research (Nash, 1988). Because of this advantage, RS technology has been widely used in many fields of geoscience (Khan and Mahmood, 2008; Klokonik et al., 2008; Worden et al., 2008; Jensen, 2009; Hugenholtz et al., 2012; Stavrakou et al., 2012; Wu et al., 2012; Kokhanovsky, 2013; Heglin et al., 2014; Panet et al., 2014; Reuter et al., 2014; Kokhanovsky et al., 2015; Tomasi et al., 2015; Bai et al., 2016; Pelpowski et al., 2016; Rajendran et al., 2014; Wu et al., 2016; Xu et al., 2016; Behrenfeld et al., 2017). In particular, satellite constellations is a major breakthrough in Earth and geospatial science, such as GEOScan (Dyrd et al., 2013).

Third, geoscience data can be processed and stored using computer technology with high efficiency. During that period, i.e., the 1970s to 2000s, computer mapping and spatial databases were regarded as two of the most important developments in geoscience research (Tomlinson, 1987; Gütting, 1994). Furthermore, computer-based modeling and simulation also promoted geoscience research (Kolditz et al., 2012; Jordan et al., 2014), such as the use of Geographic Information Systems (Napieralski et al., 2007).

However, this kind of approach still requires further development due to the lack of several critical features, including unified
management under the same type of spatiotemporal coordination, online processing of heterogeneous data sources, and a high degree of integration. One typical problem is how to efficiently use several different sensors on the same target (Wu et al., 2016; Li et al., 2017a; Li et al., 2017b). This question requires scientific modeling of mass geoscience sensors, effective collaboration methods, and reliable interconnected mechanisms. Therefore, the evolution of geoscience research has not stopped.

2.3. Precursors to the GSW-based paradigm

2.3.1. Digital earth

In 1998, Al Gore, the vice president of the USA, presented a vision for Digital Earth (Gore, 1998) as follows: “A new wave of technological innovation is allowing us to capture, store, process and display an unprecedented amount of information about our planet and a wide variety of environmental and cultural phenomena. Much of this information will be "georeferenced" - that is, it will refer to some specific place on the Earth's surface.” He emphasized that the integration of multiple sources of data and having the full range of data at our fingertips were the fundamental basis for understanding our planet in the 21st century. Other similar visions include the Infrastructure for Spatial Information in the European Community (INSPIRE) (Maguire and Longley, 2005), the Global Monitoring for Environment and Security (GMES) (Harris and Browning, 2003), and the Global Earth Observation System of Systems (GEOSS) (Lautenbacher, 2005; Lautenbacher, 2006).

With the discussion and implementation of Digital Earth, a consensus was reached that Digital Earth was a key to improving Earth systems science (Chen and van Genderen, 2008; Guo et al., 2010; Craglia et al., 2012). Many technical and scientific questions were raised in realizing Digital Earth (Goodchild et al., 2012). The most vital component was the integration and display of all earth data, including crowd-sourcing (Chen et al., 2017). One typical implementation is Google Earth by Google Inc. for displaying and conducting simple analysis on multidimensional geoscience data (Whitmeyer, 2012). Digital Earth can be regarded as a more comprehensive, fundamental, and vivid spatial database.

2.3.2. Geospatial web service

Motivated by the need for online geo-analysis capability, data sharing, and function interoperability, geoscientists have creatively introduced the open geospatial web service into the research domain (Di et al., 2005; Morris, 2006; Zhao et al., 2007). This innovation is more related to the implementation of Digital Earth. To achieve this goal, many online geoscience data-sharing projects have been established; for example, Giovanni, the Goddard Earth Sciences Data and Information Services Center (GES DISC) Interactive Online Visualization and Analysis Infrastructure, has provided researchers with advanced capabilities to perform online data exploration and analysis with observational data from NASA Earth observation satellites (Acker and Leptoukh, 2007). Other similar projects include the Federal Geographic Data Committee Sharing Policy Statements, New York State Information Technology Policy, and Ogemaw County Data Sharing Agreement. To realize these goals, geoscience-related companies and organizations have also developed numerous products, applications, and standards. For example, Open Geospatial Consortium (OGC) developed Geography Markup Language, Web Map Service, Web Feature Service, Web Coverage Service, and Web Processing Service (Peng and Zhang, 2004). Environmental Systems Research Institute (ESRI) provided ArcIMS/ArcGIS Server, ArcGIS Engine, and ArcGIS Online (Price, 2008). These developments now benefit from multi-sensor data that can be integrated and shared on a unified standardized web platform (Tu and Abdelguerfi, 2006; Amirian and Alesheikh, 2008). However, these geospatial web services/softwares concentrate on archived data instead of real time data. Moreover, there is also a lack of intelligent collaboration of multi-sensor and focusing services. Therefore, geospatial web service is only one of the necessary and basic techniques of GSW.

2.3.3. Sensor web

The concept of a sensor web was first proposed by the Jet Propulsion Laboratory (JPL) of National Aeronautics and Space Administration (NASA) back in 1999 (Delin et al., 1999). A sensor web consists of a system of intra-communicating, spatially distributed sensor pods that can be deployed to monitor and explore new environments (Delin et al., 1999). It is in the sharing of information among the sensors wherein the sensor web differs from traditional sensor networks (Delin, 2002). This global sharing of information leads to sensor synergism by
permitting intelligent resource management on the web and by allowing for self-modifying behavior based on environmental factors and internal web diagnostics (Teillet, 2010). Overall, the most distinguishing feature of sensor web is that it is a complex adaptive system and is organized as a network of open sensor resources on the Internet (Zyl et al., 2009).

Sensor webs are often confused with precursor sensor networks. However, the unique feature of a sensor web is that the information gathered by one pod is shared and used by other pods (Delin, 2002). In contrast, sensor networks merely gather data and information from a particular pod and such a network does not influence the behavior of another pod. Thus, sensor networks only collect data while sensor webs can react and modify their behavior on the basis of collected data (Delin and Jackson, 2001).

To implement a sensor web, OGC proposed Sensor Web Enablement (SWE) 1.0 and 2.0, which include several important international information model standards (e.g., Observations and Measurements (O&M), SensorML, and Transducer Markup Language (TML)) and service model standards (e.g., Sensor Observation Service (SOS), Sensor Planning Service (SPS), Sensor Event Service (SES), and Web Notification Service (WNS)) (Bröring et al., 2011a; OGC, 2011). The service model defines the interface of sensor-related web services and the information model comprising those standards, which address the specifications of the data formats (Chu and Buyya, 2007). Fig. 2 illustrates a typical sensor web service combination scenario using SOS, SES, WNS and SPS for multi-sensor-based environmental monitoring. The typical implementation of sensor webs includes work by 52° North, GeoSurf, and ArcGIS GeoEvent (Bröring et al., 2009b).

It has been suggested that increasingly pervasive sensor networks and sensor webs will revolutionize earth systems science, such as remote sensing did in the 1970s (Hart and Martinez, 2006; Zyl et al., 2009). However, the evolution of sensor web is not complete. The subsequent GSW is a remarkable new stage of sensor web evolution that truly realizes multi-sensor dynamic resource management, intelligent event perception, on-demand observation, online data processing, and focusing web services (Chen et al., 2014a).

2.3.4. Live geography

Real-time/Near real-time observations and instant processing are fantastic visions for geoscience research (McCloskey and Nalbant, 2009). One typical concept is Live Geography, which aims to combine live measurement data with historical data sources in an open standards-based infrastructure using server-side processing mechanisms. The system architecture is composed of layers of loosely coupled and service-oriented building blocks, as described in Resch et al. (2009). In this way, data integration can be decoupled from the analysis and visualization components, allowing for flexible and dynamic service chaining. To fulfill real-time data needs and alert requirements, this concept also incorporates an event-based push mechanism. This live geography can be viewed as a subset of an initial stage of GSW.

2.4. GSW-based paradigm

GSW is an advanced stage sensor web, which not only inherits the advantages of sensor webs but also possesses a greater capability for data processing and information services (Di, 2007; Di et al., 2010). In other words, the emphasis of sensor webs is mainly monitoring (Mandl et al., 2008a), whereas GSW provides a closed-loop solution for geoscience research, i.e., from monitoring, to processing, and to service (Resch et al., 2010), as shown in Fig. 3. This revolution was predicted by Hart and Martinez (2006), wherein the cyberinfrastructure of high-level data integration will bring about another major change. GSW research is now defined as a cyber-physical spatiotemporal information infrastructure for new geoscience research within unified physical and cyberspace. Therefore, GSW is not only a multidimensional monitoring network but also a real-time online analysis tool. In other words, it acts as a “bridge” between geoscience observation/phenomenon and geoscience information/knowledge/decision-making. Therefore, this approach is inherently different from real-time data publishing by individuals.
As shown in Fig. 3, there are three kinds of webs within GSW from the connotation: (1) In systems science, GSW is an integrated and collaborative sensing web. GSW connects diverse geoscience sensors or virtual sensors to form a self-organized observation system. Every single sensor can obtain geo-data alone, while mass sensors can also focus on one geo-object and share data among themselves; (2) In computing science, GSW is a scalable geoprocessing web. GSW utilizes web-based processing resources for geoscience computing. According to different scales and granularities, GSW adopts different computing resources. (3) In applied sciences, GSW is a WWW-based composable service web. GSW assembles multiple web services with standardized interfaces to accomplish one geoscience application, and also gather data/information from multiple web service for end-users.

Based on the above conceptual analyses, we found that the definition of GSW had rich connotations; while it was also evolving along with our understanding and applications. At the present stage, GSW makes full use of the advantages of multiple Earth sensing, computing and servicing resources, and serves as a WWW-based standardized cyber-physical infrastructure for geoscience research and application. Therefore, it has the following eight features:

1. GSW contains diverse web-based sensors. Regardless of satellite sensors, airborne sensors, ground sensors, underground sensors, or ocean sensors, GSW manages and accesses them using the standardized web interfaces. Regardless of what physical sensors are in the real world, or virtual sensors with standardized interfaces are available (e.g., numerical simulation system), GSW utilizes them equally.
2. All resources in GSW are sharable. GSW is a WWW-based distributed system, all sensing, computing, and service resources in GSW are equipped with standard interfaces for updating and reuse.
3. All services in GSW are interoperable. Users or third-party developers can access GSW services online.
4. GSW is a dynamic and real-time system. GSW can plan observation tasks according to previously obtained real-time data, and can then process data and publish information instantaneously. This feature is quite different from an archival data based geoscience system. Therefore, GSW is ideal for studying fast developing geo-events (e.g., forest fires, tsunami, or geohazards), which require time-continuous tracking observations.
5. GSW is an autonomous system. On the one hand, GSW can be reconfigured with user commands; On the other hand, GSW can adjust itself according to dynamic sensor deployment (e.g., sensor failure or sensor added) or the dynamic environment.
6. GSW is an extensible system. Every sensing node in GSW is a relatively independent mini system, while multiple sensing nodes can also collaborate with each other by using a standard interface to accomplish a single complex task. GSW also allows the addition of new sensing nodes with standard interfaces to expand its capability.
7. GSW is a flexible system. These sensing resources can be deployed at any specific location, and they can collaborate in different combinations for different research objects.
8. GSW has basic intelligence. This feature is represented by three aspects: GSW adjusts itself according to the target and the environment; GSW optimizes the data flow according to pre-defined task requirements and scientific objects; and GSW directly provides user-required and fused information, other than raw data.

3. Key methods of GSW

As a cyber-physical infrastructure, GSW stands between the real world and the end-user. In other words, GSW provides a complete solution for geoscience data gathering, processing, management, and service. This section reviewed four key methods of GSW, including the following: integrated management of heterogeneous sensing resources, collaborative observation based on multiple platforms, scalable processing and fusion of multi-source data, and focusing web services for researchers, government, and the public. Their roles and relationships are illustrated in Fig. 4.

3.1. Integrated management of heterogeneous sensing resources

Sensing resources are direct data sources for geoscience research; however, these sensing resources are not always the same. Sensors can be orbital, airborne, terrestrial or marine. They can be mobile or fixed. They can even be virtual sensors (Di, 2007; Chen et al., 2011a; Du et al., 2016). For example, Chen et al. (2009a) proposed an extensible sensor data adapter to find, store, and manage sensor data from various live sensors, sensor models, and simulation systems. Liang et al. (2010a) and Liang and Huang (2013) identified a long tail in the sensor web,
demonstrating most of the sensors ("missing sensors" or "dark sensors") in the tail have not been fully used. Without scientific modeling and management of these diverse sensing resources, geoscientists will not be able to utilize them efficiently. Therefore, the first key method of GSW is the integrated management of these heterogeneous sensing resources (Botts and Robin, 2007). To achieve this goal, GSW first provides two technologies: a gateway model and a unified information model.

The first is a gateway model between the sensors and WWW, which is responsible for transforming the sensor with its data from the physical world to the WWW. Typical gateway models include the Sensor Interface Descriptors (SID) (Broering et al., 2010a; Bröring et al., 2011c), Sensor Bus (Broering et al., 2010b), SensorAdapter (Liu et al., 2012), SIXTH middleware architecture (O’Hare et al., 2012), RESTful-based approach (Rouached et al., 2012; Yu and Liu, 2015), and semantically enabled approach (Bröring et al., 2011b; Gray et al., 2011). For example, Gray et al. (2011) described a semantic sensor web architecture for integrating multiple heterogeneous datasets, including live and historical sensor data, databases, and map layers. The architecture provided mechanisms for discovering datasets, defining integrated views of them, continuously receiving data in real-time, and visualizing them on screen and interacting with the data.

The second unified information model provides information mapping of these sensing resources. OGC has proposed SensorML as a standard sensor information model, which can use XML Schema to describe sensor systems and processes and provide information needed for the discovery of sensors, location of sensor observations, processing of low-level sensor observations, and listing of taskable properties (Bröring et al., 2011a; OGC, 2011). Based on these factors, a sharable and interoperable Satellite Sensor Information (SSI) model (Hu et al., 2013) and a Sensor Capability Representation model (SCRM) (Fan et al., 2015) have been proposed for more complete modeling. Fig. 5 demonstrates the SCRM modeling approach using meta-modeling architecture from bottom to top, including instances, model, meta-model, and meta-meta-model layers. This model describes a sensor based on the following five sub-capabilities: computational, transmission, observation, energy endurance, and environmental adaptability capacity (Fan et al., 2015). In particular, the observations of sub-capability can further be described by the breadth, depth, and frequency capabilities.

Based on the above model, GSW provides SOS web services to expose the sensor data interface (Chen et al., 2009a; OGC, 2011). Huang and Liang (2014) also proposed a sensor data mediator solution to define an SOS entity data model for OData (SOS-OData) to bridge these two standards. Even the uncertainties in observation can also be integrated into GSW (Stasch et al., 2012b). The aggregation of observations occurrence in space and time is needed to handle the differences in their spatiotemporal coverage and resolution, so a spatiotemporal aggregation using a geoprocessing web was also proposed (Stasch et al., 2012a).

To manage these sensing resources at a high level, a sensor web heterogeneous node meta-model was proposed (Chen et al., 2014b). Henson et al. (2009a) also argued that by providing a common model, the Observations and Measurements (O&M) facilitated syntax-level integration, but lacked the ability to facilitate semantic-level integration. This inability can cause problems with interoperability between disparate sensor networks that may have subtle variations in their sensing methods. An ontological representation of time series observations could provide a more expressive model and resolve problems of semantic-level interoperability of sensor networks on a Semantic Sensor Web (Bröring et al., 2009a). Fig. 6 demonstrates a snapshot of the GeoSensorManager, which is a typical sensing management platform in GSW developed by geoscientists from Wuhan University. Users can register a sensing resource in GeoSensorManager based on the above-proposed information model and based on retrieval data from the SOS service.

With the above methods, GSW greatly enhances the usability, interoperability, flexibility, and sharing of geoscience resources. In other words, by using web-ready sensors, geoscience data and information
3.2. Collaborative observation based on multiple platforms

Earth science research relies on multiple observations of a complex system (Reid et al., 2010; Liu et al., 2013). Early in 2002, Teillet et al. (2002) foresaw integrated Earth sensing, which connectively used both remote and in-situ sensing. At present, there are plenty of studies on multiple satellite sensor collaborations. Some of them focus on data collaboration based on certain inverse models or certain complex geo-events (Kim and Hogue, 2012; Hao et al., 2015; Wanders et al., 2015; Hollands and Dierking, 2016). In other words, multi-sensor data represent multiple parameters in an inverse model or in a complex geo-event. Other researchers have utilized multi-sensor data for long-term analyses or complex geo-event analyses (Behling et al., 2016; Zhang et al., 2017a; Zhang et al., 2017b). However, collaboration during the data capture period is missing (Zhang et al., 2014), which indicates that we usually only collaboratively use archived multi-sensor, other than actively producing collaborative data by using satellite and in-situ sensors. Observation with satellite sensors is quite different from direct observation with in situ sensors. Main differences between them are the requirement of advanced planning for satellite observation, and a certain degree of uncertainty in future observations. Therefore, we found that previous collaborative observation studies were limited due to the capabilities of satellite sensors and the orchestration required during data capturing (Zhang et al., 2014). In nature, this limitation was derived from the disconnection between heterogeneous sensors before the appearance of GSW.

Based on the above analyses, the first achievement of GSW is the evaluation of satellite sensors before an operation for collaboration. A Sensor Static Capability Index (SSCI) (Chen et al., 2015a) and a Dynamic Observation Capability Index (DOCI) (Chen and Zhang, 2014) have been proposed as two novel methods. Based on dozens of physical parameters, SSCI is able to assess the static observing capability of Spaceborne EO optical sensors with respect to a category of EO tasks or the stages of a category of EO tasks and to classify them into different clusters. Whereas DOCI is not only based on the physical parameters of sensors but also the specific task requirements. More specifically, the DOCI model consists of five sub-capabilities: spatiotemporal covering capabilities, thematic observation capability, environmental capability, attribute capability, and quality capability. Therefore, the DOCI method is more relevant to the sensor capability in future observation phases.

With an understanding of sensor capability, NASA, German Remote Sensing Data Center (DFD) and German Aerospace Center (DLR), National Cheng Kung University, and Wuhan University conducted several experiments to realize collaborative observation in GSW. In 2004 and 2006, NASA tested a multi-sensor collaboration method based on MODIS, in situ sensors, Advanced Land Imager (ALI) and Hyperion on EO-1 (Mandl, 2004; Chien et al., 2006). High-frequency MODIS measurements were used to locate terrestrial events, such as forest fires, and triggered high-resolution ALI or Hyperion sensors. In this experiment, EO-1 was driven by low-resolution MODIS to ensure that observations could be made in a timely manner and to obtain more focused images. Liu et al. (2009) also proved that combining MODIS and Formosat-2 based on optimization during the sensor planning stage could be used to rapidly locate fire points during wildfires. In 2013, DFD and DLR proposed a MODIS and TerraSAR-X collaboration method to identify and monitor the evolution of floods at an early stage (Martinis et al., 2013). Time efficiency was achieved by using daily acquired MODIS data to optimize the time-critical, on-demand programming of the high-resolution SAR acquisitions for detailed flood monitoring. Similarly, MODIS and RADARSAT were also combined to obtain instant flood detection and real-time service (Lacava et al.,...
Zhang et al. (2014) also demonstrated that the combination of long-term observation, high temporal, and high spatial resolutions sensors was valuable for soil moisture anomaly monitoring.

To realize a more organized and comprehensive collaboration, on the one hand, a Multi-Agent System and physics-based modeling were introduced into GSW (Suri et al., 2007; Bai et al., 2010; Moghaddam et al., 2010). On the other hand, an event-driven monitoring mechanism was also proposed (Fan et al., 2013). Furthermore, an existing database can also be integrated for on-demand retrieval (Chen et al., 2011b). Besides that, it was also found scientific and social data could work together in GSW (Yue et al., 2015a). GSW takes a human-as-sensor perspective and treats the social data as a special kind of sensor data that can be mined and used for event detection in the sensor web environment. Metadata requirements for sensor interoperation and synergy were analyzed as well (Di et al., 2009).

In particular, the collaboration between satellite and in situ sensors is a hot topic in GSW. It was in 2010 that the satellite measurements were firstly guided by in-situ observations and model simulations and forecasted in real time without manual intervention (Howe et al., 2010). Satellite and in situ sensor collaboration was also found to be useful in geoscience data reconstruction (Zhang and Chen, 2016; Xing et al., 2017).

At present, quite a few of the collaborative observation experiments were conducted in a simulation/theoretical way due to operation permissions. While experiments by NASA and DFD have successfully demonstrated the flexibility and value of collaborative observation in the Earth sciences (Mandl et al., 2008a; Moe et al., 2008; Martinis et al., 2013). Multiple observation platforms were interconnected within GSW (i.e., WWW) for instant cooperation. More recently, a GSW-based observation system was also tested in China on the ecological environment of the Yangtze River basin, as shown in Fig. 7. Yangtze River is the longest river in Asia and the third-longest in the world. There are several important but contradictory objectives: electricity generation, flood control, shipping, and ecological protection. It is challenging to reconcile this contradiction as they have different demands for water resources utilization. By Introducing GSW to Yangtze River management, full status of the Yangtze River has been fully obtained by satellite, airborne and ground sensors with multiple scales and purposes.

This new collaborative observation approach has fundamental differences with a traditional ground sensor network-based observation, as it promotes the monitoring and analyzing level from the local to the larger scale. With the other three methods in GSW, two new improvements had been achieved: (1) Periodical scour-silt was promoted to real time scour-silt. From 2013 to 2015 flood seasons, more water (800 million, 1.6 billion, and 1.8 billion cubic meters respectively) was stored in advance than before; (2) It used to take about 1 h to gather and send the hydrologic data to Yangtze River flood control center and national flood control center. But now, it only needs 20 to 30 min. These improvements demonstrated the collaborative observation method in GSW was particularly suitable for complex geo-event
3.3. Scalable processing and the fusion of multi-sourced data

Observation data vary within GSW due to various observation capabilities. In 2010, Teillet (2010) found that there was still a lack of multi-sensor integration, fusion and assimilation methods in the sensor web environment. Without powerful processing, the potential of multi-source data will not be fully exploited (Blum and Liu, 2005; Dong et al., 2009; Zhang, 2010). There have been several achievements in establishing a scalable processing and fusion method for multi-source data. Scalability indicates the scale and volume of processing capability are flexible and on demand. This feature benefits from the loosely coupled online processing function in GSW.

Current fusion methods are aimed at a certain purpose and lack unified fusion using spatiotemporal-spectral features at the same time (Shen et al., 2015). Moreover, most current methods are used for two kinds of sensor data fusion, other than multiple sensor data fusion for GSW. To overcome this limit, an integrated framework for spatio-temporal–spectral fusion was recently proposed (Shen et al., 2016). This framework achieved integrated fusion of multisource observations to obtain high spatiotemporal–spectral resolution images, without limitations on the number of remote-sensing sensors.

Point-surface fusion is a new scientific question that has emerged in GSW due to the collaboration of in situ sensors and satellite sensors (Zhang and Chen, 2016). There have been studies of fusion based on linear regression, multiple linear regression, and semi-empirical models. Recently, several new point-surface fusion methods based on station measurements and satellite observations have been proposed. The first category is fusion based on geostatistics, whereas the second is based on a neural network model. For example, Zhang and Chen (2016) presented a Satellite and In-situ Sensor Collaborated Reconstruction (SICR) method, which performed better than the reconstruction results only based on in-situ or satellite sensor data, as shown in Fig. 8. The Universal Image Quality Index (UIQI) by SICR is much better than other conventional methods. Then, the SICR method was upgraded using the neural network for establishing complex and highly variable relationships between the in-situ observations and remote-sensing of soil moisture (Xing et al., 2017).

In the implementation of scalable processing and fusion in sensor webs, geoprocessing web workflow is a key technology. Geoscience research relies on complex and multiple processing models, simulation models, and prediction models. The GSW method utilizes web processing together with SWE services to realize automatic and interoperable processing (Yu et al., 2008; Zhao et al., 2012). A typical geoprocessing web is GeoPW (Yue et al., 2010). Geoprocessing services are often embedded in scientific workflows and executed in a workflow engine. Sun et al. (2012a) proposed a task-oriented architecture for web geoprocessing systems (called GeoPWTManager), which leveraged web services and workflow technologies to design and execute tasks and to monitor and visualize the execution of tasks. The approach facilitates the expression of user requirements, and allows for the monitoring of task execution, and more importantly, hides the complexity of technical details. Chen et al. (2012) also proposed a method for constructing a SensorML process chain-based geoprocessing e-Science workflow. One typical tool to construct a geoprocessing workflow is Geo/ModelBuilder (Yue et al., 2015b), which is able to integrate interoperable sensors, geoprocessing services, and OpenMI-compliant model components into workflows. In this way, sensors, data, geoprocessing functions, and models could be integrated in a flexible, reusable, interoperable, and user-friendly way (Yue et al., 2015b). More advantages of this technique have been demonstrated in wildfire detection (Chen et al.,

Fig. 7. A GSW-based collaborative observation system for environmental monitoring of the Yangtze River basin.
3.4. Focusing web services for researchers and the public

A focusing web service is a methodology in GSW that is used to deliver and show the most valuable geoscience information to the most needed researchers and the public through a series of object-oriented intelligent web services (Yue et al., 2015c), other than conventional Service-Oriented Architecture (SOA). It represents the significant change from “3A geoservices” (i.e., Anywhere, Anytime, Anyone) to “4R geoservices” (i.e., Right place, Right time, Right person, Right Information).

The first achievement of focusing web services is an active feature of GSW service. To achieve it, a geo-event was first decomposed into several simple tasks based on ontology (Sun et al., 2012b). Then, collaborative observation was applied to retrieve the results of each task. After that, information required by the user was extracted and processed. Finally, the results were pushed to the subscribed users. For example, Auynirunrongkool et al. (2012) proposed an automatic instant time flood-detection approach consisting of a data retrieval service, flood Sensor Observation Service (SOS), and flood detection Web Processing Service (WPS) under a sensor web environment. MODIS data were used to overview a wide area, while RADARsat data were triggered to classify a flood area. The proposed framework, using the Transactional Web Coverage Service (WCS-T) for instant flood detection, processed dynamic real-time remote-sensing observations and generated instant flood maps for end users. Fan et al. (2013) also proposed an event-driven active on-demand data service method for forest fire monitoring. The results indicated that the proposed method could achieve desired data acquisition by subscribers in the shortest possible time. Moreover, this active feature does not even require human intervention throughout the process (Chien et al., 2006).

The second achievement is the semantic registry and discovery mechanism of GSW (Chen et al., 2009b; Corcho and García-Castro, 2010; Koubarakis and Kyzirakos, 2010). The addition of semantics in GSW makes it more intelligent (Sheth et al., 2008; Babitski et al., 2009). For example, Sheth et al. (2008) used the Semantic Sensor Web framework to enable complex queries of weather data. Jirka et al. (2009) introduced a framework for web-based discovery of sensors and sensor services. Furthermore, an ebXML Registry Information Model (ebRIM) sensor registry service was proposed, which was capable of handling the dynamic properties of sensors as well as related metadata formats and harvesting mechanisms (Zhai et al., 2012; Durbha et al., 2010).
also proposed a hybrid-matching algorithm, which exploited the instance data in ontology concepts to realize mapping between different applications. This algorithm consisted of syntactic standardization of metadata through open-standards-based sensor web components and enriching of the syntactical terms using semantics by means of conceptualizing them through ontology-based modeling.

Furthermore, the integration of GSW services with traditional geo-service is worth describing. For example, a Sensor Observation Service (SOS) chaining Web Feature Service (WFS) method was proposed to integrate geographical reference observation data collected with a hydrological Sensor Web into a virtual globe (Zheng et al., 2013). This method hid the complexity of a series of information and service models in the sensor web realm to enable the integration of heterogeneously distributed hydrological data sources into a Spatial Data Infrastructure (SDI) for end hydrological users.

In constructing better focusing web service, the quality of service in GSW was also evaluated (McFerren et al., 2009; Poorazizi et al., 2012) and, more importantly, upgraded, such as with Cascading SOS (SOS-X) (Havlik et al., 2009), SemSOS (Henson et al., 2009b), and TinySOS (Jazayeri et al., 2012).

To demonstrate the focusing service pattern of GSW, the Chinese National Disaster Reduction System of Systems (CNDRSS) was proposed as a geographically distributed, interoperable, independent, and flexible web system (Li et al., 2014). As shown in Fig. 9, CNDRSS automatically and sequentially supports disaster-related monitoring and the integration of resources from corresponding ministries using relevant templates for the workflow, resources collection, and composition of the services. Templates are not only be registered and constantly improved and updated as resources in CNDRSS Common Infrastructure but will also be customized and generated through the application layer.

### 4. Prototypes and applications of GSW

#### 4.1. GSW prototypes

Table 1 demonstrates a comparison among several typical GSW prototypes. Some prototypes are initial testbeds, whereas others are relatively mature. Sensor web simulation tools (Mekni and Moulin, 2008) have been excluded here. The following aspects have been taken into consideration: affiliation, sensor type, data model, modeling, accessing, service chain and integration with GIS. These prototypes are listed alphabetically and reviewed as follows.

The open source initiative, 52° North, has launched a series of projects to realize GSW since 2006, including six service projects, seven client projects, and four incubator projects. Nearly all implementations of GSW standards were provided, and they were the most widely used source codes in the GSW community (Kraak et al., 2005; Chen et al., 2011a; Zhang et al., 2012; Chen et al., 2015b). However, due to the design purpose, most of these projects aimed at in situ sensors and lacked a service chain. The 52°North initiative only provides a basic implementation of a single GSW service for all-purpose secondary development, instead of a complete prototype for a certain geoscience field.

GeoCENS allows users to maneuver a sensor web browser within a 3D virtual globe or on a 2D base map to discover, visualize, access and share heterogeneous and ubiquitous sensing resources, as well as relevant information (Liang et al., 2010b; Liang and Huang, 2013). In particular, GeoCENS consists of five core components, including a sensor web service engine, sensor web browser, semantic layer service, online social network, and recommendation engine. GeoCENS is innovative and unique as a social network-based sensor web platform. GeoCENS harvests the sensor web users’ interactions and activities in order to build innovative sensor web applications. With the embedded social network infrastructure, GeoCENS is able to build a geospatial folksonomy for the sensor web. This folksonomy recommends relevant sensor web resources to a user according to the collective intelligence of

![Fig. 9. Focusing service flow program within CNDRSS.](image_url)
Table 1

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Sensor type</th>
<th>Data model</th>
<th>Modeling</th>
<th>Accessing</th>
<th>Service chain</th>
<th>Integrated with GIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>52°North</td>
<td>University of Calgary</td>
<td>Satellite, aerial, in situ</td>
<td>OGC standards compliant</td>
<td>Manual</td>
<td>Manual</td>
<td>Yes</td>
<td>Two-dimensional GIS</td>
</tr>
<tr>
<td>NOSA</td>
<td>The University of Melbourne</td>
<td>Satellite, aerial, in situ</td>
<td>OGC standards compliant</td>
<td>Manual</td>
<td>Semi-automatic</td>
<td>Yes</td>
<td>Two-dimensional GIS</td>
</tr>
<tr>
<td>SANY</td>
<td>Europe Union F6</td>
<td>Satellite, aerial, in situ</td>
<td>OGC standards compliant</td>
<td>Manual</td>
<td>Semi-automatic</td>
<td>Yes</td>
<td>Two-dimensional GIS</td>
</tr>
<tr>
<td>Sensapp</td>
<td>SINTEF Norway</td>
<td>In situ sensor</td>
<td>OGC standards compliant</td>
<td>Manual</td>
<td>Manual</td>
<td>Yes</td>
<td>Two-dimensional GIS</td>
</tr>
</tbody>
</table>

4.2. Geoscience applications based on GSW

In the last ten years, GSW has been applied in a number of geoscience disciplines. As shown in Tables 2 to 4, we took a closer look at three main geoscience applications in environmental, hydrological, and natural disaster management. We found that the GSW approach served as a powerful cyber-physical spatiotemporal information infrastructure compared to a traditional manpower-based experimental approach (Di et al., 2010; Chen et al., 2013a). Not only were diverse geoscience data integrated but they were also provided in a standardized way. Not only were these data processed but they were also processed online and in real-time. This improvement allows researchers...
to have a better “connection” with the environment they study (Hart and Martinez, 2006). These studies demonstrate the remarkable improvements that could be obtained from adopting GSW in terms of efficiency, interoperability, and even ideology.

5. Challenges and perspectives

As described above, GSW has successfully played a fundamental role in geoscience research as cyber-physical infrastructure (Moe et al., 2008). To further enhance its capability, mainstream GSW research must integrate with other state-of-art technologies, thereby providing

Table 2
A review of current applications of GSW in environment research.

<table>
<thead>
<tr>
<th>No.</th>
<th>Theme</th>
<th>References</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Environmental data management</td>
<td>Conover et al. (2010)</td>
<td>The proposed method serves as the basis for a complete near-real-time environmental data management system, including sensor systems, data repositories, and registries of both sensors and observations.</td>
</tr>
<tr>
<td>2</td>
<td>Environmental data management</td>
<td>Gong et al. (2015)</td>
<td>Integrating a real-time GIS data model and sensor web service platform is an effective method to manage environmental data under the Geospatial Service Web framework.</td>
</tr>
<tr>
<td>3</td>
<td>Vegetation productivity</td>
<td>Kooistra et al. (2009)</td>
<td>A sensor web based approach combines earth observations and in situ sensor data to derive near real-time vegetation productivity products. Interoperability between the sensor data streams and connection with the information system was achieved by using open-source standards for SWE and WMS.</td>
</tr>
<tr>
<td>4</td>
<td>Soil moisture</td>
<td>Du et al. (2016)</td>
<td>GSW was used to retrieve soil moisture online and to publish soil moisture measurements as soon as receiving the BDS-R signals. This programmed mechanism was able to simplify the soil moisture retrieval process and to avoid repetitive observations, economizing time, manpower and material resources.</td>
</tr>
<tr>
<td>5</td>
<td>Soil moisture</td>
<td>Moghaddam et al. (2010)</td>
<td>The GSW technology introduced here, for integrating a physics-based modeling framework into a sensor web control system to achieve a dynamic and sparse sampling strategy, is fundamentally new.</td>
</tr>
<tr>
<td>6</td>
<td>Air quality</td>
<td>Stasch et al. (2012a)</td>
<td>Our case study indicates that the Spatiotemporal Aggregation Service allows for flexible integration of aggregation processes within SOSs and, thus, in the sensor web and geoprocessing web.</td>
</tr>
</tbody>
</table>
more complete and powerful capabilities to geoscience research. In the following four sub-sections, we discussed the most contributive and new technologies that would be necessary to integrate into GSW. Some of these integrations have generated preliminary results, whereas others have not been tested yet.

### 5.1. Integration with model web for sophisticated geoprocessing

Similar to the sensor interoperability realized by sensor webs, model interoperability was motivated by a model web. Model Web is defined as an open-ended distributed system of interoperable computer models and databases with machine services (Geller and Turner, 2007; Geller and Melton, 2008). Though GSW can provide online processing functions, its capability is still limited when handling geoscience domain models. For example, GSW is able to conduct statistical analysis, interpolation, and processing chains for precipitation or soil moisture analysis. However, GSW cannot predict future trends in land surface conditions. To solve this problem, we need both the Earth modeling study (Canepa and Builtjes, 2017) and Model Web technology. A model web utilizes the Model as a Service (MaaS) approach, relying on web services to run sophisticated models, thereby making their outputs more accessible, fostering interoperability, and working toward a larger vision of systems with independent but interacting models (Geller and Melton, 2008). The basic principles of a model web are as follows: open access, minimal barriers to entry, a service-driven approach, and scalability (Nativi et al., 2013).

Several emerging technologies could greatly facilitate the convergence of a model web, though each has limitations. One is the Earth Science Markup Language (ESML), which, along with other markup languages such as the Geography Markup Language and Ecology

---

### Table 3

<table>
<thead>
<tr>
<th>No.</th>
<th>Theme</th>
<th>References</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Water resource management</td>
<td>Guru et al. (2008)</td>
<td>GSW provides a research platform for developing next-generation hydrological and water resource management tools</td>
</tr>
<tr>
<td>2</td>
<td>Hydrological observations</td>
<td>Zheng et al. (2013)</td>
<td>The proposed approach allows for the integration of SOS servers into legacy applications that have a higher degree of availability within many spatial data infrastructure arrangements.</td>
</tr>
<tr>
<td>3</td>
<td>Ocean observations</td>
<td>Howe et al. (2010)</td>
<td>Our ocean-observing smart sensor web presented herein is composed of both mobile and fixed underwater in-situ ocean sensing assets and Earth Observing System satellite sensors, providing larger-scale sensing.</td>
</tr>
<tr>
<td>4</td>
<td>Ocean observations</td>
<td>Jiang et al. (2015)</td>
<td>A standards-based system can be built to access sensors and marine instruments distributed globally using common web browsers for monitoring the environment and oceanic conditions in addition to marine sensor data on the web; this framework of the marine sensor web can also play an important role in many other domain information integrations.</td>
</tr>
<tr>
<td>5</td>
<td>Ocean observations</td>
<td>Bermudez et al. (2009)</td>
<td>GSW contains the necessary components to represent time-series observations, not only from sensors but also from sensor systems.</td>
</tr>
<tr>
<td>6</td>
<td>Coast observations</td>
<td>Durbha et al. (2010)</td>
<td>For better interoperability among different buoy systems and enhanced data sharing among different organizations and clients, this research adopted a standards-based interoperable framework for coastal buoy data.</td>
</tr>
<tr>
<td>8</td>
<td>Water pollution</td>
<td>Hayes et al. (2009)</td>
<td>GSW will ultimately enable us to dynamically monitor and track the quality of our environment at multiple locations. The application of sensor web technology to environmental sensing will eventually result in the realization of the ‘adaptive environment’ concept.</td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>No.</th>
<th>Theme</th>
<th>References</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Disaster management</td>
<td>Wang and Yuan (2010)</td>
<td>In disaster monitoring, sensor webs play a significant role. A High-mobility Emergent Airship Monitoring System for multipurpose disaster management is needed, as a disaster may bring negative effects to the sensor web.</td>
</tr>
<tr>
<td>2</td>
<td>Disaster management</td>
<td>Hu and Chen (2011)</td>
<td>GSW was applied to change the status quo of a passive and time-lag response mode in past disaster emergency processing.</td>
</tr>
<tr>
<td>3</td>
<td>Earthquake</td>
<td>Zambrano et al. (2015)</td>
<td>We propose the use of a Sensor Observation Service and smartphones as gateways to transmit information from their embedded sensors, such as an accelerometer.</td>
</tr>
<tr>
<td>4</td>
<td>Volcano activity</td>
<td>Davies et al. (2006)</td>
<td>The fully automated process allows for rapid acquisition and transmission—typically within 48 h, though theoretically possible within 2–3 h—of data products containing the most useful data content.</td>
</tr>
<tr>
<td>5</td>
<td>Volcano monitoring</td>
<td>Song et al. (2008)</td>
<td>The Optimized Autonomous Space-In situ Sensor-web has two-way communication capability between ground and space assets, using both space and ground data for optimal allocation of limited power and bandwidth resources on the ground, and use smart management of competing demands for limited space assets. It will also enable scalability and seamless infusion of future space and in situ assets into a sensor web.</td>
</tr>
<tr>
<td>6</td>
<td>Forest fire</td>
<td>Mandl et al. (2008b)</td>
<td>In disaster monitoring, sensor webs play a significant role. A High-mobility Emergent Airship Monitoring System for multipurpose disaster management is needed, as a disaster may bring negative effects to the sensor web.</td>
</tr>
<tr>
<td>7</td>
<td>Flood</td>
<td>Delin et al. (2005)</td>
<td>The results provide an excellent opportunity to develop a mechanism for the study of flood dynamics in a controlled and well-instrumented environment. Sensor webs have great potential to change our way of monitoring and understanding hydrological processes on Earth and beyond.</td>
</tr>
<tr>
<td>8</td>
<td>Flood</td>
<td>Kustul et al. (2012)</td>
<td>GSW-based systems provide access to real-time products on rainfall estimates and flood potential forecasts and alerts. These alerts are used to trigger satellite observations. Flood maps can be generated within 24–48 h after a trigger is alerted.</td>
</tr>
<tr>
<td>9</td>
<td>Flood</td>
<td>Asurinrandronkool et al. (2012)</td>
<td>An automatic instant on-time flood detection approach, consisting of flood sensor observation service (SOS), flood detection web processing service (WPS) under the sensor web environment, is presented to generate dynamic real-time flood maps.</td>
</tr>
</tbody>
</table>
Markup Language, can act as a translator among data formats (Ramachandran et al., 2004; Ramachandran et al., 2005). Another new technology, the Earth System Modeling Framework (ESMF), can simplify the interface between ecological models and large, supercomputer-based, gridded, global climate models (Hill et al., 2004; Collins et al., 2005). Furthermore, the NASA Earth Exchange (NEX) is another good approach, particularly for larger models, models with massive outputs, or for modelers that have only limited hardware resources (Nemani et al., 2011). Recently, the GSW community has been preparing for integration with Model Web. For example, a CyberConnector has been proposed for making Earth observation data easily accessible and usable by various Earth Science Models (ESMs) (Di et al., 2016).

Model Web will increase model access and sharing, facilitate modeler-modeler and interdisciplinary interaction, and reduce re-invention. While it is difficult to quantify the costs and benefits, Model Web will provide a loosely coupled framework to make the reuse of existing services easier, thereby helping to reduce long-term modeling costs. This will not only realize more efficient use of limited model development resources but also increase the number of users, thereby resulting in more feedback to model developers and speeding up model development (Nativi et al., 2013).

To achieve this goal, the GSW community needs to collaborate with the geoscience community more closely in the future. As long as enough geoscience models have been transformed into a web-based and standardized web service, geoscientists will enjoy using these “plug-and-play” geo-models for more sophisticated geoprocessing.

5.2. Integration with humans for pervasive sensing

Human-operated sensing (human observations or humans using carry-on instruments) is a new type of sensing that is being used today and will be extensively used in the future (Teixeira et al., 2010; Doran et al., 2013). Mobile phones, social media, and volunteered geographic information (De Longueville et al., 2010; Schade et al., 2013) are typical human-operated sensing approaches. The integration of GSW with humans can further enhance the observation capability of GSW (Jürrens et al., 2009; Boulos et al., 2011). It can connect all observations conducted by all individuals in the community with mobile communication technology. It can also broadcast and share potentially interesting observations about certain individuals and obtain their feedbacks. In other words, ubiquitous sensing has become increasingly integrated into people’s living environment and in scientific research; everyone on this planet is a sensor.

By utilizing human-sensing capability, some limitations of GSW can be derived. In particular, five commonly required spatiotemporal properties can be obtained: namely, presence, count, location, track and identity (Teixeira et al., 2010). For example, in a study of geology and geomorphology, experienced experts can recognize the type of rock, the feature of a mountain, and the special soil features that may be indiscernible with GSW. A human sensor network incorporated into geo-physical models, together with satellite observations and sensor measurements, has been proposed for oil spill predictions (Aulov and Halem, 2012). Improved estimates of model parameters were obtained, such as the rates of oil spilling, couplings between surface winds and ocean currents, diffusion coefficients, and other model parameters.

The challenges of integration with human sensing are also obvious. The first is the unstructured observation data from human-operated sensing (Sheth, 2009; Huang and Xu, 2014). A data cleaning service may be needed before human-operated sensing results can be further used (Qian et al., 2009). The second challenge is the uncertainty of human observations. This uncertainty is derived from the relatively simple carry-on sensor device, or qualitative observation by humans themselves (Moser, 2005; Patt and Dessai, 2005). Bias in a human sensor web should be properly treated before using. Furthermore, these technical challenges and privacy problems should first be addressed (Li and Goodchild, 2013).

Complete pervasive sensing will be achieved only when human-operated sensing has been integrated into GSW in the future. At that time, geoscience data will be greatly enriched, and many new geoscience studies will be possible in the Anthropocene.

5.3. Integration with IoT for high-quality performance and data mining

With the increase in sensing resources in the GSW (physical, virtual, and human sensing), there will be increasing demands on computation and service resources. High concurrent requests and real-time services also depend on the hardware capability. Furthermore, geoscience models often demand high-performance computing. To overcome these challenges, GSW requires integration with Internet of Things (IoT) in the near future. Two technologies in IoT will be particularly useful for GSW, cloud computing and big data mining.

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction (Hayes, 2008; Dikaiakos et al., 2009; Mell and Grance, 2011). This technology is promising for achieving a dynamic data/service burden balance of GSW (Kussul et al., 2012). The reported high-performance based GSW applications include numerical weather prediction that is computationally intensive, flood applications that require a fast response to emergencies, and a biodiversity assessment that requires analysis and integration of large volumes of data in order to derive a final product (Kussul et al., 2009).

Future cloud-computing based GSW applications will include new visualization and interactive systems, real-time simulation and access, cloud management, spatiotemporal optimization, and global collaboration. United States Geological Survey (USGS) (NRC, 2012a) and National Science Foundation (NSF) (NRC, 2012b) also identified the typical geoscience fields enabled by cloud computing in the future, including early Earth, energy and mineral sciences, climate science, ecology, disaster management, and human and environmental health sciences.

To some extent, the concept of cloud computing is quite similar to GSW. Generally, cloud computing is a virtualization of computing hardware, whereas GSW is a virtualization of sensing hardware. Yue et al. (2013) provided a comparative analysis of GSW geoprocessing in cloud computing platforms – Microsoft Windows Azure and the Google App Engine. The analysis compared their differences in the data storage, architecture model, and development environment based on the experience developing geoprocessing services in the two cloud computing platforms. In fact, there is a consensus that using cloud computing together with GSW is able to address geoscience and Digital Earth needs within the context of an integrated Earth system (Yang et al., 2013).

Accompanying cloud computing, big data in mathematical geology and quantitative geoscience is attracting more and more attentions. Some researchers call it Big Earth Data, which has the potential to advance the in-depth development of the Earth sciences and bring about more exciting scientific discoveries (Guo et al., 2016). Compared with big data in network science or economic fields, geoscience big data has the following three features: high dimensions, high complexity, and high uncertainty (Guo et al., 2014). On the one hand, GSW produces big data, which concerns large-volume, complex, growing data sets with multiple, autonomous sources. On the other hand, GSW handles big data (Wagemann et al., 2017). As geoscience research focuses on the spatial and temporal analysis of complex geological processes, modeling theory in complex nonlinear systems, and decision support, some data analysis can be conducted using the processing capability of GSW, and some complex computation can be realized with Model Web. However, more hidden stories may only be found by big data mining (Labrinidis and Jagadish, 2012; Wu et al., 2014). Based on big data
mining, geoscience research will have a more powerful tool for analysis. The use of every piece of data from GSW in big data mining will provide solid and unprecedented input for geoscience research. Some initial work have been conducted. For example, Steed et al. (2013) proposed a big geoscience analytics system, called the Exploratory Data analysis ENvironment (EDEN), with specific applications for the analysis of complex Earth system simulation data sets. EDEN represents the type of interactive visual analysis tools that are necessary to transform data into insight, thereby improving critical comprehension of Earth systems processes.

As a state-of-art information technique, IoT will probably be the first to be integrated with GSW in the near future. This improvement is also critical for establishing a solid basis for integration with Model Web and human operators.

5.4. Integration with AI for smart geoscience research

Artificial Intelligence (AI) is intelligence exhibited by machines (Russell et al., 1995). From Deep Blue (Campbell et al., 2002) to AlphaGo (Gibney, 2016), AI has demonstrated exciting capabilities for reasoning, recognition, and prediction. Now, GSW is capable of sensing and analyzing, whereas AI can further add cognition and reasoning abilities, similar to our brains. For example, once an earthquake event is detected by GSW, the impact area and intensity can be further analyzed by GSW; however, AI can learn from previous earthquake events and the current situation, thereby determining the most effective solution for a rescue plan. Though there have been some achievements in AI-based geoscience research (Kisi et al., 2012; Ghasemi et al., 2014; Nourani et al., 2014; Shahin, 2016), AI-based GSW is still in its infancy. Limited discussions have concentrated on Multi-Agent Systems (MAS) (Suri et al., 2007; O’Hare et al., 2012).

In recent months, the vision for the Earth Observation Brain (EOB) was proposed by Li et al. (2017a, 2017b), which was an intelligent Earth observation system that simulated the sensing and cognition process of a human brain. In the future, EOB can observe the when, where, and what changes of an object to provide the right information to the right people at the right time and right place (Li et al., 2017a). Globally, all kinds of users will obtain related geospatial data, information and knowledge in real time through EOB (Li et al., 2017b). With the integration of AI into GSW, this vision will be realized. At that time, GSW is not only defined as cyber-physical spatiotemporal information infrastructure for geoscience research but also as an automatic operation center for all geoscience applications.

6. Conclusions

GSW is a revolutionary cyber-physical infrastructure for understanding the Earth. By its very nature, GSW provides spatiotemporal data and information in a form that is consistent with that requires for geoscience modeling and that represents a new paradigm for geoscience research. However, this significant revolution has not been fully reviewed yet, to the best knowledge of the authors. Therefore, this paper conducted a comprehensive and critical review of the GSW concept, historical role, key methods, prototypes, applications, and future challenges.

By adopting a GSW approach in geoscience research, scientists are able to enjoy the following four advantages: (1) integrated management of diverse sensing resources; (2) real-time or near real-time and spatiotemporal continuous data; (3) powerful and online big data processing and analyzing capability; and (4) web sharable data/information and interoperable functions. Therefore, the significance of GSW is to provide a cyber-physical information infrastructure for geoscience research, which is remarkably different and more efficient than the experiment-based and sensor-based paradigms. Though we foresee that the experiment-based approach will still play its unique role in geoscience research (e.g. determination of isotopic composition), with the help of GSW, increasingly more disciplines in geoscience will enjoy the benefit of real-time data, the collaboration of multi-sensors, and intelligence services. In other words, the ability to be in daily contact with the data source (via the World Wide Web) allows a researcher to have a better “connection” to the environment they study.

While GSW technologies are new and evolving, reference implementation and “cookbooks” for many of the services are freely available. Collaboration among domain scientists and information technology experts is critical to take full advantage of GSW in geoscience research. That is, a team comprising both geoscientists and software engineers will result in more scientifically viable, real-world results compared to a team of only scientists or only software engineers. In the future, by integrating GSW with Model Web, IoT, human, and AI, we will obtain a real-time deep understanding of the physical world for the first time at a large scale.

In 2006, Hart and Martinez (2006) concluded the significant role of Environmental Sensor Networks and foresaw its deep permeation in many areas of geoscience research and applications. The geoscience community has already experienced this profound revolution during the last decades. While more exciting and new changes have been brought by GSW in recent years, which is a remarkable new approach compared with Environmental Sensor Networks. At this background, we concluded the following after twelve years later:

GSW represents an unprecedented way to understand the Earth not only from its high dimensional exhibitions, but also from its complex and dynamic evolutions. It is a remarkable new stage of sensor networking because of its deep integration with state-of-the-art information technologies. Real-time data, multi-source information, high-performance processing and ubiquitous web services in GSW promote our understanding of the Earth in terms of both pattern and process. In future geoscience, GSW will continue to serve as a powerful cyber-physical spatiotemporal information infrastructure for more geoscience research and applications within unified physical and cyberspace.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

This work was support by the National Key R&D Program (no. 2017YFB0503803), Union Foundation of Ministry of Education of the People’s Republic of China (no. 6141A02022318), Creative Research Groups of Natural Science Foundation of Hubei Province of China (no. 2016CFA003), National Basic Research Program of China (no. 2011CB707100), National Natural Science Foundation of China program (no. 41171315, 41771422), the Natural Science Foundation of Hubei Province (no. 2017CBB016), Creative Research Groups of National Natural Science Foundation of China (no. 41021061), China Postdoctoral Science Foundation (no. 2017M620338, 2018T110804), and LEIESMARS Special Research Funding (2018-06).

References
