Seismology Perspectives on Integrated, Coordinated, Open, Networked (ICON) Science

Lei Li\textsuperscript{1,1}, Wing Ching Jeremy Wong\textsuperscript{2,2}, Benjamin Schwarz\textsuperscript{3,3}, and Tsz Lam Lau\textsuperscript{4,4}

\textsuperscript{1}Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring, Central South University
\textsuperscript{2}Scripps Institution of Oceanography, University of California San Diego
\textsuperscript{3}GFZ German Research Centre for Geosciences
\textsuperscript{4}The Chinese University of Hong Kong

November 30, 2022

Abstract

Seismology focuses on the study of earthquakes and associated phenomena to characterize seismic sources and Earth structure, which both are of immediate relevance to society. This article is composed of two independent views on the state of the ICON principles (Goldman et al., 2021) in seismology and reflects on the opportunities and challenges of adopting them from a different angle. Each perspective focuses on a different topic. Section 1 deals with the integration of multiscale and multidisciplinary observations, focusing on integrated and open approaches, whereas Section 2 discusses computing and open-source algorithms, reflecting coordinated, networked, and open principles. In the past century, seismology has benefited from two co-existing technological advancements - the emergence of new, more capable sensory systems and affordable and distributed computing infrastructure. Integrating multiple observations is a crucial strategy to improve the understanding of earthquake hazards. However, current efforts in making big datasets available and manageable lack coherence, which makes it challenging to implement initiatives that span different communities. Building on ongoing advancements in computing, machine learning algorithms have been revolutionizing the way of seismic data processing and interpretation. A community-driven approach to code management offers open and networked opportunities for young scholars to learn and contribute to a more sustainable approach to seismology. Investing in new sensors, more capable computing infrastructure, and open-source algorithms following the ICON principles will enable new discoveries across the Earth sciences.

Hosted file

essoar.10508630.2.docx available at https://authorea.com/users/555246/articles/608168-seismology-perspectives-on-integrated-coordinated-open-networked-icon-science
Seismology Perspectives on Integrated, Coordinated, Open, Networked (ICON) Science

Lei Li\textsuperscript{1,2}, Wing Ching Jeremy Wong\textsuperscript{3,4}, Benjamin Schwarz\textsuperscript{5}, and Tsz Lam Lau\textsuperscript{3}

\textsuperscript{1}Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring, Ministry of Education, Central South University, Changsha, China
\textsuperscript{2}School of Geosciences and Info-Physics, Central South University, Changsha, China
\textsuperscript{3}The Chinese University of Hong Kong, Hong Kong, China
\textsuperscript{4}Scripps Institution of Oceanography, University of California San Diego, California, USA
\textsuperscript{5}GFZ German Research Centre for Geosciences, Potsdam, Germany

Corresponding author: Lei Li (leileely@126.com)

Key Points:

- We comment on the current status and potential for synergy and challenges of implementing ICON principles in seismology.
- The integration of multi-parametric and multi-scale observations across disciplines benefits Earth imaging and earthquake understanding.
- High-performance computing and open-source algorithms offer networked opportunities for a broader community to contribute to seismology.
Abstract

Seismology focuses on the study of earthquakes and associated phenomena to characterize seismic sources and Earth structure, which both are of immediate relevance to society. This article is composed of two independent views on the state of the ICON principles (Goldman et al., 2021) in seismology and reflects on the opportunities and challenges of adopting them from a different angle. Each perspective focuses on a different topic. Section 1 deals with the integration of multiscale and multidisciplinary observations, focusing on integrated and open approaches, whereas Section 2 discusses computing and open-source algorithms, reflecting coordinated, networked, and open principles. In the past century, seismology has benefited from two co-existing technological advancements - the emergence of new, more capable sensory systems and affordable and distributed computing infrastructure. Integrating multiple observations is a crucial strategy to improve the understanding of earthquake hazards. However, current efforts in making big datasets available and manageable lack coherence, which makes it challenging to implement initiatives that span different communities. Building on ongoing advancements in computing, machine learning algorithms have been revolutionizing the way of seismic data processing and interpretation. A community-driven approach to code management offers open and networked opportunities for young scholars to learn and contribute to a more sustainable approach to seismology. Investing in new sensors, more capable computing infrastructure, and open-source algorithms following the ICON principles will enable new discoveries across the Earth sciences.

Plain Language Summary

Seismological observations can provide critical insights into the physical processes of the Earth’s interior and associated near-surface consequences, resulting from both natural and anthropogenic activities across spatial and temporal scales. This commentary discusses the current status and opportunities of integrated, coordinated, open, and networked (ICON) principles in seismology. As an applied discipline, seismology is highly data-dependent and inherently relies on sensing (data acquisition) and computing (data processing and modeling) technologies. Integrating data from multiple scales and domains has improved our understanding of earthquakes and is also beneficial to adjacent disciplines, such as reservoir engineering and rock mechanics.
When more open data and models are produced and integrated by coordinated acquisition systems and networked programming efforts, seismology will enable new discoveries in the Earth sciences.

1 Introduction

Seismology is an applied discipline and integrates techniques and data from physics, mathematics, informatics, mineralogy, and geology. Seismological studies involve various natural and anthropogenic activities across a wide range of spatial and temporal scales, including tectonic plate motions, volcano eruptions, hydrocarbon exploration, carbon sequestration, mining, landslides, and laboratory stimulation experiments (Stein & Wysession, 2003; Shearer, 2009). Seismological observations can provide critical insights into the physical processes of the Earth’s interior and the associated near-surface consequences (Cloetingh & Negendank, 2010; NASEM, 2020). The multidisciplinary nature and multiscale observations of seismological studies embody integration, corresponding to the ‘I’ in ICON science. There are many seismological data centers, such as the International Federation of Digital Seismograph Networks (https://www.fdsn.org/about/) and the Incorporated Research Institutions for Seismology Data Management Center (IRISDMC, https://ds.iris.edu/ds/), providing findable, accessible, interoperable, and reusable (FAIR) seismic data collected at local, regional to global scales, which generally comply to international and coordinated data format standards (e.g., SAC, MiniSEED, SEED, and SEGY). These FAIR data and consistent protocols and standards are coordinated efforts benefiting open exchange, therefore representative of the ‘C’ and ‘O’ in ICON science. Besides, there are numerous code packages and libraries openly accessible and extendable for seismological studies. This modern community-driven approach to programming, along with the improved availability of computational resources and machine learning algorithms, is a networked (‘N’ in ICON science) effort which has been significantly promoting seismology and adjacent disciplines, such as reservoir engineering and rock mechanics.
2 Integration of multiscale and multidisciplinary observations

Every breakthrough in seismology is always dependent on advancements in sensing and/or computing technology. The science of seismology was born around the 1880s, along with the invention of time-recording seismographs (W. H. K. Lee et al., 2002). The upgrade of seismic instrumentation (e.g., broad-band seismographs) since the 1930s and the advent of plate tectonics and modern computers in the 1960s enable seismologists to exploit the rich information encoded in the seismograms, determine Earth’s fine-scale internal structure, and quantify the diverse spectrum of fault slip behaviors (e.g., Peng and Gomberg, 2010). The past few decades have witnessed the arrival of large and dense arrays at regional/national scales (e.g., USAArray at http://www.usarray.org/, HiNet at https://www.hinet.bosai.go.jp/, and ChinArray at http://www.chinarraydmc.cn/) and ocean-bottom seismometers, advancing array-based analysis techniques and yielding more detailed reconstructions of seismic sources and Earth's internal structure (Rost & Thomas, 2002; Karplus & Schamndt, 2018; Cai et al., 2018; L. Li et al., 2020). More recently, rotational seismographs and fiber optic sensing technology have been pushing seismology and other related disciplines a giant step forward regarding data acquisition. Traditionally, seismology has been dealing with pure translational motions of the ground, whereas rotational motions - predicted by the theory of elasticity - could not be captured by conventional seismic sensors. Only recently, Earth's rotational seismic field has been captured through the integrated analysis of dense station networks or newly developed rotational sensors (e.g., Lee et al., 2009). Distributed acoustic sensing (DAS) or fiber-optic seismology, as an exciting example of interdisciplinary intersection which can also benefit the theory and technology of fiber-optic sensing, alleviate seismological observation bias due to limited temporal and spatial resolution by transforming permanent fiber optic cables into sensor arrays with meter-scale (and higher) resolution. Recent studies have demonstrated the reliability of fiber-optic cables, deployed under the ground or the seafloor, in delivering densely sampled strain or strain rate measurements of seismic wavefields to study phenomena in the cryosphere, marine geophysics, geodesy, and volcanology (Lindsey & Martin, 2021; Zhan et al., 2021). First field applications already suggest that fiber-optic strain sensing will allow substantial improvements in resolution and sensitivity in critical regions while remaining uniquely cost-effective (e.g., Jousset et al., 2018). The combination of sensors for translation, rotation, and
DAS provides novel information about deformation caused by seismic disturbances. However, conventional data protocols and standards for translational motions may not apply to rotational and DAS observations, requiring additional efforts to achieve full integration (‘I’) and coordination (‘C’) of these measurements. Moreover, the unprecedented station density of DAS – resulting in Terrabytes of data per day – currently inhibits open data exchange and availability and poses serious challenges to integrated studies involving these new observations. Meanwhile, enhanced acoustic emission (AE) sensors (e.g., piezoelectric ceramic transducer (PZT)) and instruments promote signal-based AE analysis, which enables in-depth analysis of slip mechanisms, rupture propagation, and expected damage under controlled conditions in the laboratory (Ishida et al., 2017; Wang et al., 2017; Brotherson, 2021).

Enhancing multiscale and multidisciplinary observations, and promoting the sharing and exchange of datasets under the ICON principles will benefit seismology and relevant disciplines, such as reservoir engineering and rock mechanics. In the past two decades, seismology shifted its focus towards (micro)earthquakes caused by human underground activity such as mining, shale gas exploitation, and geothermal energy production (Grigoli et al., 2017; Foulger et al., 2018; Schultz et al., 2020). Figure 1 illustrates how human activities can have an influence on the stress state of the Earth's crust and induce (micro)earthquakes. With advanced acquisition and processing techniques, induced microearthquakes can be used for fracture geometry delineation and reservoir geomechanical analysis, while larger-magnitude earthquakes are essential for earthquake hazard analysis (L. Li et al., 2019). Besides, laboratory experiments and numerical modeling can help characterize the thermal-fluid-solid coupling during fluid injection. Multiscale and multiphysics investigations combining elastic and electromagnetic wavefields, deformation, and temperature will facilitate reservoir characterization and underlying physical mechanism understanding, thereby aiding risk assessments and possible mitigation efforts. This integrated approach to confronting an important challenge in seismology is synonymous with integrated (I), coordinated (C), and networked (N) science. However, the conflicting interests in industry may result in decreased transparency and open (O) exchange. This will hinder the cooperation and mutual benefits between scientific and industrial communities, and more regulations and means of communication are still in demand to strengthen this important interface.
Figure 1. Induced seismicity monitoring associated with various industrial activities involving fluid injection. Induced (micro)earthquakes can not only aid reservoir characterization and guide subsurface operations but also provide crucial insight into how peoples’ safety and the protection of local infrastructure can be ensured. On the one hand, spatial and temporal high-resolution sampling of seismic wavefields can be achieved by surface and downhole arrays, and DAS cables; on the other hand, the combination of multidisciplinary measurements, including seismic, electromagnetic, temperature, and ground deformation monitoring can also better constrain data processing and interpretation.

One of the most important and long-standing research questions in seismology remains: can we forecast earthquakes more accurately (AAAS, 2021)? A potential and feasible strategy is integrating multiscale seismological and even multidisciplinary observations. Acoustic emissions (AE) from laboratory earthquakes exhibit striking similarities to natural earthquakes. In laboratory experiments, scale-independent earthquake properties can be measured with a high spatial and temporal resolution. However, there exist only a few FAIR databases of local and smaller scale earthquakes. Seismic datasets for exploration purposes are barely open and freely shared due to commercial concerns or risks, and laboratory AE datasets are also not extensively and systematically managed. Although seismology has experienced great progress during the past decades, the depth and breadth of intersection between seismology and adjacent
disciplines (e.g., geophysics and geochemistry) are not yet sufficient. For earthquake seismology, scientists have attempted to interpret seismic signals based on structural geology and tectonics since the 1960s. Besides geological and physical properties of faults and rocks, more attention should be paid to lithological and mineralogical features to characterize the Earth’s interior as a dynamic system. To obtain a predictive understanding of earthquake mechanisms, future multidisciplinary and networked (N) research should also integrate (I) fault/fracture complexities, mineralogical phase changes, and thermal-fluid-solid coupling into earthquake models. This again highlights the integral importance of the ICON principles in seismological research. At the exploration scale, research and applications combining these two categories of seismic methodologies are thriving in recent years (Berkhout and Vershuur, 2011), largely due to the popularity of dense monitoring arrays. However, the in-depth integration of seismic and other geophysical approaches (e.g., electromagnetic measurements) is still rare. For indoor experiments, multiphysics observations, including mechanical, seismic, and optical measurements, are combined to uncover the dynamic process of earthquake nucleation and fracture propagation. In-situ laboratory scale experiments can help reveal the site effects of seismic responses and bridge the inherent scale gaps of seismological studies. How to better integrate multiscale observations and multidisciplinary (including geological, geophysical, geochemical) processes is a major challenge and task for an improved understanding of earthquake hazards.

3 Advancements in computing and open-source algorithms

Like in other fields, the rapid growth of computing infrastructure and the ever-increasing amount of data implies a shift towards big-data analysis and advanced numerical techniques in seismology. This new brand of computational seismology heavily builds on numerical source and wavefield simulations and the (joint) inversion of massive datasets. In general, computational seismology broadly covers the following aspects:

- Numerical simulations to model earthquake rupture dynamics and related
hazards

- Data mining of the seismological recordings to extract useful information
- Data management and code development

Graphics Processing Unit (GPU) programming and cloud computing are enabling new integrated workflows that allow for extensive and realistic modeling of earthquake processes. In the past decades, it was only possible to construct simplified models and boundary conditions to elucidate distinctive mechanisms of seismic sources. Peta and exa-scale computing facilities promise to allow a detailed description of the fault geometry, tsunami-earthquake coupled simulations, large scale numerical solutions of the wave equation for signals with high frequencies, and the realistic modeling of interactions with the surrounding medium (Igel, 2017). These full-scale numerical simulations enabled by high-performance computing (HPC) infrastructures are critical yet have not routinely been implemented for the rapid response and assessments of cascading earthquake hazards (Hori et al., 2018). Taking the 2018 Palu-Sulawesi earthquake-tsunami event as an example, a preliminary tsunami warning was canceled soon after the earthquake report of a strike-slip event, leading to escalated damage by the surprising tsunami (Ulrich et al., 2019). A variety of initial studies have been proposed and shared on social media immediately, reporting on the confirmation of supershear earthquake rupture, complex fault geometries, and surface deformation from InSAR measurements (Lacassin et al., 2020). These scattered resources are crucial in effective rapid damage evaluation and response to earthquake and tsunami hazard, albeit being difficult to manage and coordinate. The recently established Centre of Excellence for Exascale in Solid Earth (ChEESE, https://cheese-coe.eu) by the European Union and the Earthquake Simulation project (EQSIM) by the US Department of Energy’s Exascale Computing Project (ECP) aim to coordinate the sparse and preliminary resources and enable the urgent supercomputing earthquake hazard simulation (de la Puente, et al., 2020, McCallen et al., 2021). With the emergence of peta- and exa-scale computing facilities, more efforts such as the piloting ChEESE and EQSIM initiatives are needed to coordinate to facilitate a networked (N) exchange and hazard response for the broader scientific community and the public sphere.

Apart from numerical modeling, observations are the primary requisites to extract
hidden signals of different dynamics of the Earth. Enabled by the open availability and interoperability of large databases, recent advances in machine learning (ML) are paving the way towards automating critical yet often time-consuming tasks in seismology. Examples of successfully leveraging ML include the detection and picking of seismic arrivals (Perol, et al., 2018; Ross et al., 2018; Zhu & Beroza 2018; Mousavi et al., 2020). Deep neural-network architectures originally developed for computer vision or speech recognition were able to reliably extract patterns and predict phase arrivals in seismic time-series data. Aside from mere pattern recognition, ML was shown to also lend itself well to a broad range of other tasks, including data augmentation, solving partial differential equations (PDEs), and computing synthetics using neural networks in an efficient way (Bergen et al., 2019; Morra et al., 2021). Combining array- and ML-based techniques, we are now witnessing a new era of Real-time Intelligent Array Seismology (RIAS) (Li et al., 2021). Fortunately, powerful ML libraries are open and easily accessible and are generally accompanied by detailed instructions and tutorials. This new era of machine learning application requires extensive knowledge that is not part of the traditional earth science or science education. The publicly available resources offer practical experience for students and researchers to join and facilitate the rapid advances of ML in seismology. Therefore, it remains crucial to continue sharing traceable, reproducible, and open-source codes in concordance with the FAIR data policy (Wilkinson et al., 2016). Physics-informed neural networks are an emerging trend next to the data-driven ML approaches to further explore and interpret physical processes (Raissi et al., 2019). With additional physical constraints on ML algorithms, fewer training samples and computational efforts are required to obtain a more generalized inference than traditional methods can provide. The data-centric approach of ML will make the ICON-FAIR principles ever more important in computational seismology: more open datasets and models are required and will need to be integrated, which are ideally produced with coordinated acquisition systems and mutually beneficial (i.e., networked) programming efforts. Inversion, as another essential processing technique in seismology, is implicitly related to ICON principles through state-of-the-art computation resources and algorithms/codes. The community-led effort Collaborative Seismic Earth Model (Afanasiev et al., 2016) aims at recovering scale-consistent properties of the Earth interior and is a prime example of a networked and open approach to the field.
Advances in programming language and hardware design, mathematics and computer science require a flexible approach to scientific software development. Code development in seismology often used to be restricted to isolated and specialized research groups. However, open-source code management platforms like GitHub or GitLab, offer an opportunity for the seismological community as well as other interested individuals to develop and share codes on regular and/or on demand bases. This paradigm shift in code development from closed research groups to open community-driven and individual users provides equal opportunities to participate in collaborative studies. These current trends in computational seismology are in accordance with the core ICON principles: the computational, mathematical, and physical sciences are integrated with seismology, open data and algorithms/software are generated with coordinated and networked efforts by a broader community. One of the renowned examples is ObsPy - A Python Toolbox for Seismology/Seismological Observatories that provides a framework for basic processing of seismological data (Beyreuther et al., 2010). The community-driven ObsPy package has rapidly evolved with a large group of over 90 code contributors and numerous other commentators. The package has gained popularity within the seismological community with more than 50 seismological analysis packages built upon the ObsPy framework. Besides, the open and networked style of code management also facilitates education and training for young scholars, which in turn ensures the sustainable development of seismology. We realize that it is impossible to cover all community efforts relating to ICON principles in this short commentary. There are many prominent and coordinated community initiatives in place that keep having a profound impact on the field. Prominent examples include the Incorporated Research Institutions for Seismology (IRIS, https://www.iris.edu/), the Southern California Earthquake Center (SCEC, https://www.scec.org/), and more recent efforts such as the initiation of the fully community-driven diamond open-access scientific journal Seismica (www.seismica.org). Following the ICON principles, a coordinated integration of multidisciplinary and multiscale measurements in conjunction with the increasing availability of distributed computational resources and openly developed scientific software will improve reproducibility and sustainability of seismology and enable new discoveries within Earth sciences and beyond.
Acknowledgements

We thank the editor Kristy Tiampo for handling this paper, the reviewer A.-A. Gabriel and two anonymous reviewers for their helpful comments. L.L. acted as a facilitator and point of contact with the special collection organizing team. We appreciate Dirk Gajewski, Zhigang Peng, Sujata R. Emani, and James C. Stegen for reviewing and proofreading the manuscript. The work is sponsored by the National Natural Science Foundation of China (Grant No. 42004115), Hunan Provincial Natural Science Foundation of China (Grant No. 2019JJ50762).

References


