High-performance computing for understanding seismic sources

*Satoshi Ide[1], Takane Hori[2], Ichimura Tsuyoshi[3]

1. Department of Earth an Planetary Science, University of Tokyo, 2. JAMSTEC, 3. Earthquake Research Institute, University of Tokyo

Slip inversion using seismic waveform data has become a popular method to investigate the source process of large earthquakes. This is fundamental information to understand physical process related to earthquake rupture and to assess the possibility of future earthquakes. However, the results usually include large errors, which is mainly due to the limitation of our ability to compute seismic wave field between the source and stations. There are actually two problems: the lack of information about underground structure and technical difficulty to solve elasto-dynamic equations in complex structures. The latter can be tackled with state-of-art high-performance computing.

In practical data analyses of earthquake rupture process, the availability of data is limited to avoid the source of large model errors. We often assume, without guarantee, that near-field strong-motion records are well approximated by 1-D layered structures, and that PREM and the ray theory are sufficient for teleseismic data in the hypocentral distance range from 30˚ to 100˚. Data between 2˚ and 30˚ were not used, usually. These data are plenty and have sufficient signal-to-noise ratio, but the wave propagation is complex and information about structure is insufficient in this intermediate range.

A breakthrough to this problem might be provided by the combination of high-performance computing and high-resolution structural information. Several studies have attempted to compute seismic wave fields with complex structural information. Tsuboi et al. (2016) calculated seismic waves up to 1.2 s, using a global 3D heterogeneous model. Ichimura et al. (2016) computed deformation of a space of about 3000 km, with a realistic underground structure around Japan, and suggested that similar computation for seismic field is also possible. Seismic waveforms using realistic Earth structure will become popular in near future.

Although seismic wave field in global heterogeneous Earth is currently too expensive, significant improvement is expected just by introducing realistic shape of subducting slab and seafloor topography. The shape of subducting slab is clearly resolved around Japan. However, in some subduction zones, the uncertainty of hypocenter locations is too large to image slab shape. Near-filed seismic stations are rare, though many stations are available in the intermediate range. In such a region, theoretical waveforms calculated considering slab and see-floor topography may be useful.

In medium scale, and in a well-studied region with dense observation networks, like Japan, more detailed study including a complex underground structure may be meaningful. It would be helpful to know how the result of seismic tomography and receiver function analyses improve the resolution of the seismic source imaging. Starting from the low-frequency limit, i.e. static deformation, we will be able to model complete seismic wave field around Japan, which will reduce the uncertainty of seismic source models.

Very informative, but challenging data for seismic source imaging are the records of ocean bottom seismograms, which have been installed in many places around Japan. In addition to complexity due to seafloor topography and water layer, appropriate modeling of thick sediment layers is necessary to explain complicated observed seismograms. This was a realistic problem when an earthquake of Mw 6.5 occurred in the Kumano-Nada region on April 4, 2016. Because of our inability to model seismic waves in this region, we could not distinguish whether the earthquake was inter- or intra-plate earthquake, immediately after the earthquake. In this region, active tectonic tremors and slow slip events were observed by DONET1, and underground structure is relatively well-known. If accurate theoretical waves are available using high-performance computing, these are helpful for studying both large disastrous earthquakes and

diverse slow earthquakes.

Keywords: Earthquake source process, Theoretical waves, subduction zones

Moment tensor inversion of 2016 Southeast Off Mie earthquake via numerical simulation using a three-dimensional velocity structure model

*Shunsuke Takemura[1], Takeshi Kimura[1], Katsuhiko Shiomi[1], Hisahiko Kubo[1], Tatsuhiko Saito[1]

1. National Research Institute for Earth Science and Disaster Resilience

On April 1, 2016, Mw 5.8 earthquake (2016 Southeast Off Mie earthquake) occurred near the epicenter of the 1944 Tonankai earthquake. To investigate seismic activity around the Tonankai area, source mechanism of the 2016 Off Mie earthquake should be required. However, since seawater, accretionary prism (low-velocity sediments) and subducting Philippine Sea plate exist beneath the epicenter region, it is difficult to obtain accurate source mechanism via conventional one-dimensional analysis (Nakamura et al. 2015; Takemura et al. 2016). Thus, in this study, we conduct moment tensor (MT) inversion using Green's function via finite-difference method (FDM) simulations of seismic wave propagation in a 3D heterogeneous velocity structure model (hereafter, simply called "3D Green's function"). Recent developments of computer system and calculation technique enable us to evaluate 3D Green's functions practically.

The model of 3D simulation and technical details are same as in Takemura et al. (2016). The 3D heterogeneous velocity structure, including topography, sedimentary layer, crust and subducting Philippine Sea plate, is referred from the Japan Integrated Velocity Structure Model (JIVSM; Koketsu et al., 2012). By using displacement waveforms for periods of 30-100 s, we estimate an MT solution of this earthquake. The result with minimum variance reduction is the optimal solution with source mechanism and centroid depth.

The optimal result is characterized by a low-angle dipping faulting at a depth of 11 km, where the upper surface of the Philippine Sea plate exists closely. Obtained result reproduced not only long-period displacement waveforms but also polarity of short-period (~ 2 s) *P* waves. Observed *P*-first arrivals at land stations show apparent velocity of approximately 7 km with up polarizations, which propagate along oceanic Moho of the Philippine Sea Plate

Since our FDM simulations evaluated seismic wave propagation for periods longer than 1.4 s, calculation of one element moment tensor at a certain depth requires 1.3 TB of computer memory and a wall-clock time of 1.5 hours by parallel computing using 256 nodes (1,024 cores) of the Earth Simulator.

Global seismic wave computation on the K computer

*Seiji Tsuboi[1], Kazuto Ando[2], Takayuki Miyoshi[3], Daniel Peter[4], Dimitri Komatitsch[5], Jeroen Tromp[6]

1. JAMSTEC, Center for Earth Information Science and Technology, 2. RIKEN,RIST, 3. Univ. Tokyo, Earthquake Research Institute, 4. KAUST, 5. Aix-Marseille University, 6. Princeton University

We present high-performance simulations of global seismic wave propagation with an unprecedented accuracy of 1.2 seconds seismic period for a realistic three-dimensional Earth model by using the Spectral-Element Method on the K computer. Our seismic simulations use a total of 665.2 billion grid points and resolve
1.8 trillion degrees of freedom. To realize these large-scale computations, we optimize a widely used community software code to efficiently address all hardware parallelization, especially thread-level parallelization to solve the bottleneck of memory usage for coarse-grained parallelization. The new code exhibits excellent strong scaling for the time stepping loop, i.e. parallel efficiency on 82,134 nodes relative to 36,504 nodes is 99.54%. Sustained performance of these computations on the K computer is 1.24 petaflops, which is 11.84% of its peak performance. The obtained seismograms with an accuracy of 1.2 seconds for the entire globe should help us to better understand rupture mechanisms of devastating earthquakes.

Keywords: seismic wave ppropagation, spectral element method, K computer

Modeling crustal deformation with mechanical models and geodetic data

*Jun'ichi Fukuda[1], Kaj M. Johnson[2]

1. Earthquake Research Institute, University of Tokyo, 2. Indiana University

Over the past two decades, development of dense and continuous geodetic networks makes it possible to capture transient crustal deformation at high resolution in both space and time. A primary example of geodetically observed transient deformation in plate boundary zones is postseismic deformation following large earthquakes. Postseismic deformation observed at the surface reflects mechanical properties of faults, crust, and upper mantle. Thus mechanical models that are designed to fit the observed postseismic deformation may provide insights into these mechanical properties at depth.

Two primary processes that are responsible for postseismic deformation are viscoelastic relaxation of coseismic stress changes in the upper mantle and afterslip on the plate interface surrounding the coseismic rupture. Viscoelastic relaxation is governed by the rheological properties of the crust and upper mantle, while afterslip is governed by frictional properties of the plate interface. Both of the processes are initiated by stress changes due to coseismic slip. Therefore, geodetic observations of postseismic deformation could be used to constrain the frictional properties on the plate interface and upper mantle rheology.

In this talk, we present a three-dimensional coupled model of stress-driven frictional afterslip and viscoelastic stress relaxation for postseismic deformation following the 2011 Mw9.0 Tohoku-oki earthquake. In this model, we assume that afterslip is governed by a velocity-strengthening friction law that is characterized with a friction parameter (a-b) σ . Viscoelastic relaxation of the upper mantle is modeled with a biviscous Burgers rheology that is characterized with the steady-state and transient viscosities. We calculate the evolution of afterslip and viscoelastic relaxation using an assumed coseismic slip model as the initial condition.

We examine the effects of the friction parameters, mantle viscosities, elastic thickness of the slab and upper plate, and coseismic slip distribution on the model prediction and find that these parameters significantly affect the predicted surface postseismic displacements. Postseismic deformation following the 2011 Tohoku-oki earthquake has been captured by both on-land GNSS and seafloor GPS/Acoustic networks. We thus explore the range of the model parameters that can fit the postseismic observations. At this moment, we employ a trial-and-error approach to estimate the parameters. However, given the complexity of the model and the abundance of the observations, it is difficult to completely characterize uncertainties and trade-offs between the parameters by the trial-and-error approach, although such information is critically important for better understanding of the postseismic processes. A more complete model that fully accounts for uncertainties and trade-offs between the parameters could be obtained by building posterior probability distributions of the parameters using Bayesian inverse methods. Such Bayesian methods, however, require many forward calculations and thus it would not be easy to implement those methods with our model within a realistic computation time. It is therefore essential to reduce the computational costs of the forward model simulations and Green's functions calculations, as well as to develop more efficient Bayesian methods to estimate posterior probability distributions, in order to develop more advanced models of geodetically observed postseismic deformation.

Keywords: crustal deformation, mechanical model, geodetic data

A Bayesian hierarchical model for a seismic source inversion

Amato Kasahara[1], *Yuji Yagi[2]

1. Graduate School of Life and Environmental Sciences, University of Tsukuba, 2. Faculty of Life and Environmental Sciences, University of Tsukuba

The earthquake rupture processes of earthquakes are diverse reflecting the heterogeneous distribution of fault constitutive properties and stress, and fault geometry. Kinematic source parameters constrained by data provide the information related to the fault rupture dynamics and statistical properties of rupture processes. It is also used as an input for the calculation of Coulomb stress changes, the calculation of tsunami waveforms and the estimation of slip deficit.

A source inversion analysis using spatio-temporal displacement field data is able to be formulated as a discrete linear inverse problem when the Green function and the source fault are known. However, it is difficult to calculate an accurate Green function due to the lack of the accurate Earth structure model. Effects of the uncertainty of the Green function on the result of an inverse analysis increases with the improvement of the quality and the increase of the amount of data. Some of previous studies approximated the effects of the uncertainty of the Green function by introducing a new correlated and/or uncorrelated error term, which is added to data. The approximation, however, fails to capture important characteristics of the effects of the uncertainty of the Green function such as peak shift and heavy tails of the likelihood function under the uncertainty of the Green function.

To address the issue, we propose a hierarchical Bayes model for a seismic source inversion analysis. The model is targeted to the multi-data analysis with the multi-time-window finite fault source parameterization. We assume Gaussian observation errors and a Gaussian prior distribution for model parameters. In the model, a Green function is treated as a realization of a random variable G. We approximate the posterior distribution of the model parameters by approximately marginalizing G and the hyperparameters, which control the variance of the prior distribution of the model parameters and the variance of the observation errors. The marginalization of the hyperparameters is approximated by plugging in the maximum a posteriori hyperparameters given G. The marginal likelihood function for G is approximated by the Laplace approximation to the conditional posterior distribution of the hyperparameters given G. The marginalization of G is approximated by a Monte-Carlo integration method. The precisions of the two approximations are able to be improved by increasing the amount of data and number of samples of G.

We applied the method to synthetic data for far-field vertical P-wave displacement waveforms. We set a 1-D velocity structure consisting of two layers at the source region. We set the P-wave velocity, the S-wave velocity and the density of each layer to follow uniform distributions. We also set the depth of the interface of the layers to follow a uniform distribution. We made the true velocity structure by randomly sampling the 25th percentile point or the 75th percentile point for each random variable. We drew thousands of velocity structure samples from the probability density function (PDF), and then calculated the corresponding Green functions. For reference, we also conducted a conventional inversion, which used only a reference velocity structure (50th percentile point of the PDF of the velocity structure). We found that the conventional inversion result suffers artifacts especially at the later shallow part of the rupture process, while the result with the proposed method does not suffer from the artifact. Note that the mitigation of the artifact was not possible with the simple mean-of-the-posterior-mean approach. We also

found increase of the Variance of the posterior distribution of the potency due to the marginalization of G.

The method could be a reasonable counterpart for the recent increase of Bayesian approaches with Monte-Carlo methods to study velocity structures.

Keywords: Bayesian hierarchical model, Seismic source inversion

Creating future of solid Earth science with HPC: Discussion

*Takane Hori[1]

1. R&D Center for Earthquake and Tsunami, Japan Agency for Marine-Earth Science and Technology

Due to the development in computer science and computational science, large-scale or many times forward simulations and/or inversion analyses have become available recently. In solid Earth science, large-scale seismic wave propagation and crustal deformation with high fidelity model based on high resolution observation data have been demonstrated; uncertainty in crustal deformation caused by material properties and structures can be investigated based on many-time calculations for different material properties and structures; fault slip inversion analyses for non-Gaussian error distribution, etc. Thus, we invited researchers who are facing problems in forward simulations and inversion analyses. In this talk, we will discuss how to solve such problems by the collaboration between computer & computational sciences and solid Earth science.

Keywords: high performance computing, high fidelity model, data analysis

A study on the characteristics of amplitude and dominant period of long-period ground motions via 3D finite-difference simulation: A case study for the Kanto sedimentary basin

*Kazuo Yoshimoto[1], Kengo Kajikawa[1], Ayaka Izumi[1], Kaori Shimazu[1], Shunsuke Takemura[2]

1. Graduate School of Nanobioscience, Yokohama City University, 2. National Research Institute for Earth Science and Disaster Resilience

Introduction

In the Kanto sedimentary basin, long-period ground motions (LPGMs) are frequently observed for shallow local and regional earthquakes. It is reported that, in the Kanto sedimentary basin, the predominant period of LPGMs is mainly controlled by the bedrock depth of sedimentary basin (e.g., Yoshimoto and Takemura, 2014b), but with the exception of the area around the Tokyo Bay where the elongation effect due to low-velocity sedimentary layers is observed (e.g., Kajikawa et al., 2016). It is also found that the excitation of LPGMs is very complex and varies by epicentral direction of earthquakes (e.g., Yuzawa and Nagumo, 2012). For the better seismological understanding of these observations, we conducted 3D finite-difference (FDM) simulations using simplified velocity structure models of sediment-basement system to evaluate the effect of sedimentary structure and basement shape on the characteristics of strength and predominant period of LPGMs.

Simulation method and structure models

We conducted 3D FDM simulations for investigating the characteristics of LPGMs during shallow moderate earthquakes. For the evaluation of LPGMs for periods longer than 4 s, models with a volume of $150{\times}60{\times}72$ km^3 discretized by a grid interval of 0.15 km were used for this simulation. Other technical details are same as in Takemura et al. (2015). We used three sedimentary structure models referred from VSP measurements (Yoshimoto and Takemura, 2014a): Yokohama model (Y-model), Chiba model (C-model), and Iwatsuki model (I-model). We adopted four basement topography shapes: flat model (basement depth of 3.5 km) and flat + triangle-shaped hill (height of 2.5, 3.0, and 3.5 km) models. Structure beneath the basement was a stratified layered structure referred from the JIVSM (Koketsu et al., 2012). Pure strike- and reverse-type point source models, which excites Love and Rayleigh waves, respectively, were used in our simulations.

Result of numerical experiments

(1) Characteristics of Love and Rayleigh waves

Single-peak envelope of the fundamental-mode Love wave was found in the simulation using a strike-type source. In contrast to this, for a reverse-type source, multi-peak envelope of LPGMs appeared. These findings indicate that in the case of reverse-type source, LPGMs are constructed by superposition of not only fundamental- but also higher-mode Rayleigh waves. Dominance of horizontal amplitude over vertical amplitude observed for the simulations using a reverse-type source is consistent with this interpretation.

(2) Bedrock topography dependent PGV

Peak ground velocity (PGV) of LPGMs increases with increasing height of a triangle-shaped hill on the basement. This result is interpreted by the strengthening of the Airy phase of Love waves by the constructive interference of short period surface waves that are effectively excited and propagate at shallow depths. This phenomenon may correspond to the strong excitation of Love waves at the northern and western boundary of the Kanto basin.

(3) Local amplification of LPGMs

Local variation of the PGV of LPGMs was detected during surface wave propagation thorough adjacent sediments with different velocity structures. For example, propagation of Love waves from Y-model to C-model showed an amplification of LPGMs at period longer than about 5 s, and especially at about 10 s. This result successfully explains the observed local amplification of Love waves in the western coast of the central Boso Peninsula (Kajikawa et al., 2016).

The characteristics mentioned above were also found in large-scale FDM simulations using JIVSM or SBVSM (Masuda et al., 2016 SSJ). However, these simulations could not sufficiently reproduce the observed LPGMs around the Tokyo Bay and western edge of the Kanto Basin. To overcome this problem, detailed analysis of observed LPGMs, forward modeling and estimation of sedimentary structure via waveform inversion should be required.

Keywords: Long-period ground motion, Kanto sedimentary basin, Sedimentary structure, Surface wave, Maximum amplitude, Predominant period

Quasi-static earthquake cycle simulation based on large-scale viscoelastic finite element analyses

*Ryoichiro Agata[1], Ichimura Tsuyoshi[2], Mamoru Hyodo[1], Sylvain Barbot[3], Takane Hori[1]

1. JAMSTEC, 2. Earthquake Research Institute, The University of Tokyo, 3. Nanyang Technological University

Earthquake cycle simulation is extensively studied in the field of solid earth science as a tool to explain earthquake generation processes. It is also expected to play an important role in disaster mitigation, such as generation of possible earthquake scenarios as inputs for earthquake damage estimation. An approach combining a boundary element approach based on Green's function in an elastic half space and the rate- and state-dependent friction law is widely used for this simulation (e.g. Hori 2009; Barbot et al. 2012). On the other hand, in crustal deformation computation, complex physics such as the mantle rheology and the effect of gravity are not negligible in some cases. To consider extensibility for such effects, it is desirable to develop an earthquake cycle simulation combining crustal deformation computation based on numerical simulation such as finite element (FE) method with the rate- and state-dependent friction law. This approach used to be practically difficult because of the associated computational cost, but the recent development of a fast and scalable FE solver (Ichimura et al. 2016) assuming use of supercomputers is expected to make it feasible. Therefore in this study, we seek to apply the crustal deformation computation using the viscoelastic FE analysis method developed by Ichimura et al. (2016) to earthquake cycle simulation based on the rate- and state-dependent friction law.
We use the equation of motion, the rate- and state-dependent friction law, and a slowness law as the governing equations of the earthquake cycle simulation. However, instead of computing stress changes along the fault plane by superimposing slip response function as in the previous studies, we compute them using the time history calculation of viscoelastic deformation using the FE method. For now, the time integration scheme and other components of the simulation method all follow the approach in Hyodo and Hori (2014).
We are now performing verification of the developed simulation code using a normative three dimensional problem, where a circular-shaped velocity-weakening area is set in a square-shaped fault plane. In the presentation, we will discuss the comparison of our results with those obtained using the previous methods. If possible, we would also like to discuss the change of earthquake generation process due to the introduction of viscoelasticity in the cycle simulation, as an example of the effect by mantle rheology.

Keywords: Earthquake cycle simulation, Finite element method, the rate- and state-dependent friction law, large-scale numerical simulation

Forward simulation of postseismic process after moderate and large interplate earthquakes along the Nankai Trough

*Mamoru Hyodo[1], Ryoichiro Agata[3], Ichimura Tsuyoshi[2], Takane Hori[1]

1. Earthquake and Tsunami Forecasting System Research Group R&D Center for Earthquake and Tsunami, Japan Agency for Marine-Earth Science and Technology, 2. Earthquake Research Institute, The University of Tokyo, 3. School of Engineering, The University of Tokyo

Along the Nankai trough, it has been considered that low seismic activity in the interseismic period is a typical feature there, and that the accumulated strain during the interseismic period is only released by great interplate earthquakes with recurrence intervals of 90-200 years. However, recent observations or studies show that there have occurred several M6~7-class interplate earthquakes within the region of the Nankai Trough earthquake, such as an M6-class earthquake at Kumano-nada (April 1, 2016) and an M7.6 Hyuga-nada earthquake (July, 1498).

Recent two M6~7-class earthquakes occurred within the source area of great interplate earthquakes nearby the Japanese Islands were followed by contracting consequences: the 2016 Kumano-nada earthquake above wasn't followed by the great interplate earthquake, while an M7-class interplate earthquake which occurred at off-Tohoku region in March 9, 2011 was followed by the 2011 Tohoku-oki earthquake, and was regarded as a foreshock of Tohoku-oki earthquake.

Such different consequences are caused by the different urgency of the great interplate earthquake at the timings of occurrences of these M6~7-class earthquakes. Thus, when the occurrence of great earthquake is approaching, postseismic slips after the M6~7-class earthquakes can easily propagate to the surrounding region and can trigger the subsequent great event. While, if the fault is not urgent to the great event, strong coupling on the fault may prevent postsesmic slip from propagating to the surrounding area. These differences in postseismic slip patterns strongly affect the corresponding crustal deformation observed at the earth's surface or the ocean bottom. Hence, from occurrence pattern of the crustal deformation after the M6-7 earthquake in the target source region, it might be possible to narrow down the pattern of afterslip or the subsequent scenario that can occur.

In this study, we focus on the Nankai Trough region, and firstly examine the possible propagation pattern of after slips after the hypothetical occurrence of earthquakes such as M7-class Hyuga-nanda earthquake or M6-class Off-Kumano earthquake in many Nankai Trough earthquake scenarios deduced from numerical simulations. Then, we classify the expected postseismic deformation patterns at the pre-existing observation networks depending on the propagation pattern of particular after slip. Further, we will consider whether the pre-existing network have a distinguishability of after slip propagation pattern or not. If the distinguishability in the existing network isn't enough to distinguish the postseismic slip, we will seek for the better observables or the better observation arrangements. In oder to evaluate realistic crustal deformation, we will evaluate crustal deformation associated after slips not only in a homogeneous elastic half-space but also in FE models with heterogeneous crustal structure and the configuration of bathymetry at the earth's surface.

Keywords: Nankai Trough earthquake, crustal defomation, urgency of the great earthquake

Challenge of Preparing for Careers in Big Data in Geosciences

Larry Zheng[1], *Gabriele Morra[2,1], David A. Yuen[3,1], Davin Loegering[1], Henry Tufo[4,1], Chuck Li[5,1]

1. Mc Data , Wuhan, 2. Department of Physics, University of Louisiana at Lafayette, LA, 3. Department of Earth Sciences, University of Minnesota, Minneapolis, MN , 4. Department of Computer Science, University of Colorado, Boulder, CO, 5. Wuhan Huawei Technology Co., Ltd

In the aftermath of the 2008 financial crisis we have seen the steady encroachment of Big Data into every facets of society, from finances to medical services. Students graduate lacking technological skills despite needing them in the lab and on the field. We believe that putting a stronger emphasis on programming and technology will prepare them for the demands of today's modern job market in the geosciences and to use better measurement and analysis technology.

Our curriculum in educating students needs some changes, but universities move too slow. Therefore training centers are sorely needed. For this reason, we have established Mc Data Consult Ltd., based now in Wuhan, but poised to move anywhere.

Our aims are four fold:
(1) To establish training courses at both fundamental and advanced levels, which will be taught with customized software embedded within a affordable data-analytic tool box built with (a) cheap processors such as Raspberries Pi and (b) higher-end Nvidia TX1. Students can learn and perform exercises according to their available time slots.
(2) To provide professional consulting for various Big Data challenges encountered in industries.
(3) To hold workshops and international conferences where we can mix people from various disciplines and engage them in Big Data immersion.
(4) We also see the need to prepare suitable textbooks , focusing on high-performance computing, visualization and data analytics. We maintain that Python holds the key for preparing the students in Big Data analytics.

To be sure, the big data problem is not a new paradigm for geoscience. For instance, Peter Shearer (1991) used a relatively simple 1-dimensional velocity model to stack thousands of long-period body waves, revealing two upper mantle discontinuities, which was the first successful "big data" application: the primary computing happens for data processing, not for artificial modeling. Thus, we believe that geoscientists can be prepared to adapt to the big data era once they master the modern tools: they should master an open programming language suitable for large data, such as Python, and know how to harness parallel and distributed systems. They should learn sound software engineering skills, just as a wet chemist needs to learn to wash glassware. They should learn to produce a reproducible work: all analyses should be scripted and point-and-click tools should be avoided. They should have skills in data visualization and should master the rudiments of nonparametric, computationally based statistical inference, such as permutation tests.

Keywords: Big Data, Machine Learning, High Performance Computing, Python, Education