Project Details

Name of project: A cloud-based community tool for ambient seismic noise analysis and monitoring with Julia

Project lead and contact details: Dr. Dylan Mikesell, Associate Professor of Geophysics, Phone: 208.426.1404 (dylanmikesell@boisestate.edu), Boise State University, 1910 University Drive, MS 1535, Boise, ID 83725, USA

Project partners and contact details: Incorporated Research Institutions for Seismology (IRIS), IRIS DMC, 1408 NE 45th Street, Suite 201, Seattle, WA 98105

1) Dr. Jerry A Carter, Director of Data Services, Phone: 206.547.0393, (jerry@iris.washington.edu)

2) Rob Casey, Deputy Director, Cyberinfrastructure, Phone: 206.547.0393 x117, (rob@iris.washington.edu)

Proposed start and end date: 1 August 2019 -- 29 February 2020

Budget Requested: $7,000

Budget Summary: Write instrument response deconvolution library (incorporate StationXML.jl) -- $500; Integration of DistributedArrays.jl into the Noise.jl package -- $1250; Integration a database for tracking cross correlations (need to determine optimal database type) -- $750; Web application development with AWS CodeStar (incorporate Noise.jl, SeisIO.jl, PlotSeis.jl) -- $3000; Optimization, testing, debugging, user manual, and software documentation -- $1500

Project Outline

Project description: We propose to develop a cloud-based (e.g. Amazon Web Services) software portal to process massive amounts of seismic data for 1) earthquake hazard characterization and 2) temporal monitoring of dynamic processes in the Earth's crust (e.g. landslides and volcanoes). The long term goal of this project is to transition from giga-scale seismology to peta-scale seismology through partnership with the Incorporated Research Institutions for Seismology (IRIS), the world's largest seismological data center, and development of an open-source cloud-based software application for seismological researchers. We plan to develop a Julia-based cloud-computing application that will enable anyone to process and interpret seismic data. The Julia computing language has been developed to exploit all the possible speedups afforded by modern hardware, including GPUs. It outperforms most languages in standard benchmark exercises and in the I/O tests that we ourselves have performed. The PI plans to work with a scientific computing student researcher to implement and optimize a variety of recently developed seismic data processing algorithms into a single open-source application. Our ultimate goal is to facilitate modernization of the ambient seismic noise correlation data processing workflow at scale (from GB to PB) with an application that runs on laptops, workstations, HPC centers, and in the cloud. In this seed project, we will incorporate existing open-source codes written in the Julia language into a cloud-based application. The goal of this seed project is show a proof-of-concept application that will enable a larger proposal in the future that leads to new understanding of the dynamic processes happening with in Earth. This future proposal will further modernize other aspects of data intensive processing in seismology beyond ambient noise cross correlation.

Objectives: The main objective of this seed project to create a working web application on AWS that interfaces with IRIS for seismic data download and then performs parallel seismic noise cross correlation. To achieve this, we will integrate existing Julia packages into one web application. We will build a web application using AWS's CodeStar, a cloud-based service for creating, managing, and working with software development projects on AWS. This tool enables us to interface with a Git repository on our local machines to build out new packages and integrate with existing packages.

Significance: While high-throughput algorithms and applications are being developed for many different physical science disciplines (e.g. Celeste in Astronomy), the seismic monitoring community is lagging behind. This in contrast to seismology's historical leading role in numerical modeling using HPC resources. The monitoring community has mostly moved to an open-source platform written in Python called Obspy. This tool has promoted the democratization of seismology data processing, but is particularly inefficient at handling large data sets. Moreover, the most commonly used simulation software (e.g. for wave propagation in a heterogeneous 3D Earth) is still written in Fortran90 or C++ for speed, precluding modification or innovation by more classically trained geoscientists. The proposed software targets the broad seismological community for data processing, and by gaining efficiencies using cloud-based parallel computing with Julia, we eventually plan to develop this application to enable real-time data processing to create a tool that seismologists can use to benefit society in real-time (e.g. earthquake warning). The nearest tool the seismological community currently has is MSNoise, a Python tool based on Obspy for seismic noise processing that is not parallelized, nor does it offer sophisticated preprocessing tools. It offers the bare minimum to the community. This project will demonstrate the potential of Julia and cloud-computing to the seismological community.

Impact: As seismic sensors become cheaper, more and more sensors are being installed around the world in geologically hazardous areas, and real-time data are being sent to and stored at data centers. The total global seismic data stored by IRIS is 0.5 PB (see letter of support), and a new technology called Distributed Acoustic Sensing (DAS), which re-purposes fiber optic cables as seismometers, will soon be generating PetaBytes of seismic data annually. Recently, Ajo-Franklin et al.[1] created 0.3 PetaByte of seismic data in 6 months using

[1] Ajo-Franklin, J. B., Dou, S., Lindsey, N. J., Monga, I., Tracy, C., Robertson, M., … Li, X. (2019). Distributed Acoustic Sensing Using Dark Fiber for Near-Surface Characterization and Broadband Seismic Event Detection. *Scientific Reports*, *9*(1), 1328. http://doi.org/10.1038/s41598-018-36675-8

DAS. Therefore, because of the massive amount of continuously recorded data being collected right now, seismologists are finding themselves at a breaking point, and the data are rarely fully utilized due to the exhaustive computing that it demands, even though the problem is *embarrassingly parallel*. Besides the lack of open-source high performance tools that could democratize high-level computing, few researchers have access to super-computing facilities with unlimited core hours. Thus providing this new cloud-based tool will enable scientists all over the world to solve new problems and improve our understanding of natural hazards, and potentially their precursors.

Description of key project steps and timeline:

(Aug.-Sep. 2019) We will use Joshua Jones' SeisIO.jl to access data at IRIS via web requests. SeisIO.jl is a collection of utilities for reading and downloading geophysical time series data. Designed for speed, efficiency, and ease of use, it includes web clients, readers for common seismic data formats, and fast file writers. Utility functions allow time synchronization, data merging, padding time gaps, and basic data processing. **(Sep. 2019)** We will write our own seismic instrument deconvolution code in Julia. This is straightforward and we can use existing models from Python and Fortran90 libraries. This step is necessary in order to cross correlate different types of recording instruments; in essence we remove the imprint of the instrument on the seismic data. **(Oct-Dec. 2019)** We will integrate an existing noise correlation package (Noise.jl) from Dr. Marine Denolle's PhD student Tim Clements and the PI's own MATLAB-based package into the web app. This is where we will need to decide how to parallelize the cross correlation of time series. We will likely use the DistributedArrays.jl package, but we will test other options as well (e.g. MPI.jl). (Dec. 2019) We will integrate the ASDF file format for parallel file I/O. **(Oct. 2019-Jan. 2020)** We will build these packages into a single integrated noise correlation package and then into a web application. **(Jan.-Feb. 2020)** We will test, debug, and optimize the code. A simple user interface will be created that allows the user to select stations and time periods over which to compute noise correlations. The code will provide correlation stacking options (e.g. Phase-Weight Stack), as well as a number of different correlation and pre-processing methods (e.g. frequency filter). The user will be able to download a single ASDF file that contains stacked cross correlations. Lastly, we will need to integrate a database to keep track of time series correlation progress, as well as metadata for each time series prior to stacking and writing ASDF data. This will be a bare-bones working application example that we can use to demonstrate the possibilities of a cloud-based noise correlation tool.

Description of additional funding currently supporting this work: There is no funding currently supporting this work for the PI. The PI is in contact with the Julia developers previously mentioned and has been working to discuss the process moving forward in terms of collaboration on a bigger project to build a Julia-based seismic processing toolbox, equivalent to Obspy and MSNoise in Python. Dr. Marine Denolle has been funding a postdoc to develop Noise.jl based on the PI's MATLAB code.

Outreach

What groups/audiences will be engaged in the project? Initially, the seismic monitoring community will be engaged in this project. There are earthquake and volcano observation centers in many countries around the world. Most of these groups will benefit from the results of this project, especially those groups who host their data live at IRIS.

How will you judge that project has had impact? We will judge impact by how many researchers try this new code and start to use it for regular data processing (track Forks on GitHub). If the seed project results interest the community enough that multiple groups get together to write a larger proposal, then this seed project will have made its impact. (It may be that the real impact is noticed after the larger proposal, but we still anticipate a number of groups using the software created by the seed project.)

How will you share the knowledge generated by the project? The PI will present the results at the 2021 AGU Annual Fall Meeting. This is the largest annual gathering of seismologists. We will also submit a publication to *Seismological Research Letters -- Electronic Seismologist*.

Description of *who* (agencies/individuals) should be aware of this project, i.e. potential outreach targets: The seismological community should be aware of this project: including researchers at universities, private companies, and government agencies (e.g. USGS, DoD). IRIS can inform the seismology community via the Data Services Newsletter.

Project Partners (as applicable)

Description of project partners (agencies/individuals) and their involvement: See the attached letter of support from Dr. Jerry A Carter, Director of Data Services at IRIS. We will also work with Drs. Joshua Jones and Marine Denolle. Dr. Jones has experience in Julia with web-based data access from IRIS and hardware optimization. Dr. Denolle has experience with parallel crosscorrelation computation and Julia. Dr. Denolle will be a collaborator on the larger proposal after this incubator project.

How will this project engage members of the ESIP community: Seismic monitoring is fast becoming the biggest producer and consumer of seismic data. Many ESIP community members are seismologists, and many member institutions have research seismologists within their ranks. This project applies to all of those people.