

# QuakeSim: Integrated Modeling and Analysis of Geologic and Remotely Sensed Data

Andrea Donnellan, Jay Parker, Robert Granat, Eric De Jong, Shigeru Suzuki  
Jet Propulsion Laboratory, California Institute of Technology  
4800 Oak Grove Drive  
Pasadena, CA 91109  
818-354-4737  
Andrea.Donnellan@jpl.nasa.gov

Marlon Pierce and Geoffrey Fox  
Indiana University  
2719 East 10th Street  
Bloomington IN 47408

John Rundle  
Department of Physics, University of California, Davis  
One Shields Avenue  
Davis, CA, 95616

Dennis McLeod and Rami Al-Ghanmi  
University of Southern California  
Mail Code 0781, 3651 Trousdale Parkway  
Los Angeles, CA 90089

Lisa Grant Ludwig  
University of California, Irvine  
Irvine, CA 92697

*Abstract*—The QuakeSim Project improves understanding of earthquake processes by integrating model applications and various heterogeneous data sources within a web services environment. The project focuses on the earthquake cycle and related crustal deformation. Spaceborne GPS and Interferometric Synthetic Aperture data provide information on near-term crustal deformation, while paleoseismic geologic data provide longer-term information on earthquake fault processes. These data sources are integrated into QuakeSim’s QuakeTables database and are accessible by users or various model applications. An increasing amount of UAVSAR data is being added to the QuakeTables database through a map browsable interface. Model applications can retrieve data from QuakeTables or remotely served GPS velocity data services or users can manually input parameters into the models. Pattern analysis of GPS and seismicity data has proved useful for mid-term forecasting of earthquakes and for detecting subtle changes in crustal deformation. The GPS time series analysis has also proved useful for detecting changes in processing of the data. Development of the QuakeSim computational infrastructure has benefitted greatly from having the user in the development loop. Improved visualization tools enable more efficient data exploration and understanding. Tools must provide flexibility to science users for exploring data in new ways, but also must facilitate standard, intuitive, and routine uses for end users such as emergency responders.<sup>1</sup>

## TABLE OF CONTENTS

<b>1. INTRODUCTION.....</b>	<b>1</b>
<b>2. DATA STORAGE AND ACCESS .....</b>	<b>2</b>
<b>3. COMPUTATIONAL INFRASTRUCTURE.....</b>	<b>3</b>
<b>4. INTERFACES TO EXTERNAL APPLICATIONS .....</b>	<b>4</b>
<b>5. VISUALIZATION .....</b>	<b>5</b>
<b>6. SCIENCE PROBLEMS AS A DRIVER.....</b>	<b>5</b>
<b>7. CONCLUSIONS .....</b>	<b>6</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>6</b>
<b>REFERENCES .....</b>	<b>6</b>
<b>BIOGRAPHIES .....</b>	<b>7</b>

## 1. INTRODUCTION

Earthquake science depends greatly on numerous data types spanning spatial scales from microscopic to global and timescales of fractions of seconds to millions of years. As the inadequate preparation and response to recent major earthquakes in Haiti, Chile, and Japan have shown, the field is ripe for transformation: formerly isolated groups must work more effectively with each other. Data providers need to better understand how their data are consumed and fused with other data sources by downstream geophysicists. Geophysicists must understand how to relate their work to emergency planners and responders. Experts focused on the processes of particular processes or areas of the globe must find ways to translate their knowledge to other regions and other research teams. All must be focused on identifying and tackling grand challenges that span areas of expertise.

<sup>1</sup> 978-1-4577-0557-1/12/\$26.00 ©2012 IEEE.

Collaboration alone is not enough: the field needs a common framework designed to foster the desired connections. This is especially imperative as datasets and sources grow as new spaceborne missions and ground based networks contribute to the field.

QuakeSim is a multi-source, synergistic, data-intensive computing infrastructure for modeling earthquake faults individually and as part of complex interacting systems. Remotely sensed geodetic data are integrated with models and pattern analysis applications in a rich web-based and visualization environment. The goal is to integrate heterogeneous data [Table 1] and various tools to develop models efficiently, allow rapid exploration of large datasets, and identify subtle but important features in large datasets. QuakeSim is valuable for earthquake investigations and modeling in its current state, and also serves as a prototype and nucleus for broader systems under development.

Numerous and growing online data sources from NASA, USGS, NSF, and other resources provide an exceptional opportunity to integrate varied data sources to support comprehensive efforts in data mining, analysis, simulation, and forecasting. The primary focus of QuakeSim is fault modeling and its application to earthquake forecasting, but the developed technology can support a wide array of science and engineering applications. QuakeSim development has resulted in a number of successes but has also identified a number of key challenges related to data, computational infrastructure, and model, analysis, and visualization infrastructure.

Earthquake research activities are hampered by the uncoordinated (but improving) state of current data collections, and the lack of formal modeling tools capable of ingesting multiple data types. Addressing these issues requires a comprehensive set of activities. These include 1) developing bridging services in a service-oriented architecture to integrate data from multiple sources, including interferogram, GPS position and velocity measurements, and seismicity; 2) devising a fundamental framework for model optimization through the integration of multiple data types; 3) developing cyberinfrastructure within science gateways to handle the computing requirements of the optimization framework including the need to access large datasets; 4) enabling effective data acquisition, management, and sharing - involving data semantics, provenance, version tracking, commenting, rating, etc.; and 5) developing capabilities for delivering data to downstream applications.

Key challenges identified through science analysis and QuakeSim integration include the need for more open processes, particularly in the creation of data products, and the greater integration and accountability of different groups. The solutions are partially technical and partially sociological.

**Table 1. Data Sources**

Observation	Characteristics
GPS	Position time series Velocities
UAVSAR	Line of site change
Seismicity	Magnitude and location
Fault	Location, geometry, slip parameters
InSAR	Line of site change

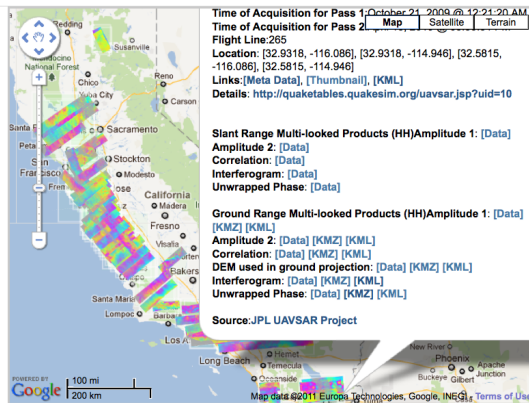
## 2. DATA STORAGE AND ACCESS

Integrating and modeling the ever-growing and increasingly multisource geodetic GPS and Interferometric Synthetic Aperture Radar (InSAR) data volumes is necessary to improve fault models. The models are used for forecasting and simulation, as well as for emergency planning and response applications. Estimates of crustal deformation from remotely sensed data are key to improving fault models. GPS data provide long-term estimates of crustal deformation of networks of sites. A time series of the daily change in position of these sites provides detailed information about temporal crustal changes. Current InSAR data products provide detailed images and spatial distribution of crustal deformation sparsely sampled in time. The planned DESDynI-R mission will provide routine high-resolution interferograms of crustal deformation [1]. This will add a significant increase in the temporal and spatial resolution of InSAR data products.

Efficiently analyzing, integrating, and modeling geodetic and geologic data requires digital storage of the data, including the fault specifications, and automated access to the data through network services. As the data sources, volumes and regions of interest grow it is necessary for applications, not just humans, to access the data for remote automated processing. The data are distributed and under the cognizance of a wide array of agencies and institutions. Developing standards through formal and informal collaborations and partnerships is key to maximizing the use of solid Earth science data. Numerous processes result in deformation of the Earth's surface. Accessing, mining, and modeling crustal data are key to understanding these processes. The potential applications of Earth surface data, such as simulations, are varied, include a large globally distributed set of users, and create archives many-fold times larger than the centers that process or store the data.

Data storage, processing, mining, and analysis challenges need to be addressed now to maximize the utility of the planned DESDynI-R mission where the design, launch, and operation represent a significant cost. The benefits of that investment will only be realized if the infrastructure is in place for investigators and users to access and interpret the data. Regardless, such infrastructure will enable greater utility of UAVSAR and international InSAR mission data.

## UAVSAR RPI Map



**Figure 1. QuakeTables database showing UAVSAR map browser and links to data products.**

Models require an increasing number of types of data to guide them. The data are of many different forms and sizes. Fault data, for example, yield information about fault geometry, slip rates, and earthquake recurrence. At the other end of the spectrum interferometric radar data tend to be in large binary image files on the order of 1 GB/image. QuakeSim applications use fault data, GPS time series and velocities, seismicity data, seismic moment tensor information, and interferometric synthetic aperture radar (InSAR) images. QuakeSim applications use data products, rather than the raw data. Other organizations process the data, turning them into products.

Understanding the origin and processing of the data products is important for assessing their quality for ingestion into models. Data products often change with time as new processing techniques or new interpretations become available. One key challenge is keeping up with improved solutions as they become available. Ideally, there is a feedback loop from modelers to data product providers that enables modelers to identify issues with the data and request reprocessing of data. This feedback loop results in greater utility of the data. Training modelers to process complex data as well is not always reasonable. Both data processing techniques and model development are so complicated that they can take careers to develop and as a result teams of people rather than individuals must contribute to the final analysis.

Data products, even for the same data type are not standardized, and are often not adapted for machine interfaces. This requires manual input, or often, at best scraping of web pages for information. While this is not the right approach, it is often the only available approach. Standardized service interfaces are needed for interfacing data with modeling and visualization tools. Data formats should be standardized through community use cases. Data product needs for earthquake science are as follows:

- All data products should be coupled with self-describing network service interfaces. A great deal of useful data and metadata about earthquakes, for

example, is bound in human-readable web pages instead of machine-readable formats (e.g., ontologies).

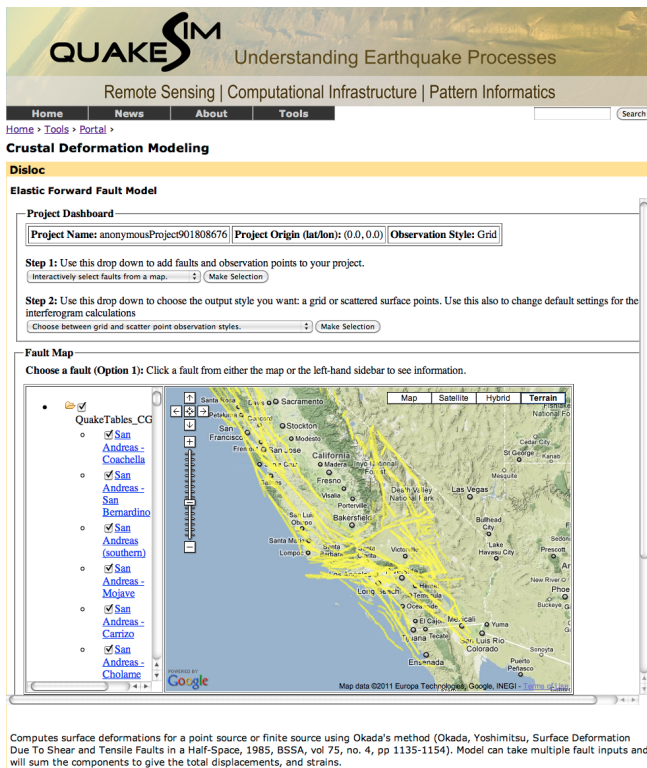
- Services should be documented, published, and discoverable.
- Services for analyzing lower level data products should also be designed with the same approach. These services generate products that may be consumed downstream.

Data presented in a map view that can be browsed eases selection of the data (Figure 1). Information about data is often encoded into long file names, and often locations over which the data are collected are encoded as station names or flight paths. Without the user having familiarity with the identification scheme it is often difficult to locate data of interest. A map view of the data makes it easier for a user to efficiently scan for data over regions of interest. Problems arise when data are collected over different time spans and overly other data in the same region, or multiple data interpretations exist. Pop-up lists, menus, or time slider bars can alleviate some of these issues. Data in the QuakeTables database can be accessed directly by applications through web service APIs as well as through visual browsing.

### 3. COMPUTATIONAL INFRASTRUCTURE

A user friendly computational infrastructure is necessary for identifying and pulling in data from numerous sources, simplifying or automating data assimilation, mining, and modeling workflow, and providing feeds and interfaces for generalized data users. QuakeSim provides a back-end for earthquake forecasting and response, crustal deformation modeling, and modeling of fluids within the crust. The scaling of compute power should occur at the back end and be transparent to the user.

QuakeSim applications require the user to do the following either in an automated manner or with user intervention: 1) select data in terms of types, time, and space; 2) subset data to relevant focus of interest; 3) move data for mining, modeling, or visualization; 4) analyze data by modeling, inverting, or data mining; 5) visualize data and results; and 6) track data and models. For small data sets or regions of interest these steps can be done manually and in fact such investigations provide excellent examples for developing workflow for larger and more complicated cases. Current data volumes and in particular those for existing or planned InSAR missions motivate the need for an end-to-end architecture in which data can be systematically analyzed, modeled, and interpreted. Automation requires interfaces between the widely distributed data sets, data products, and applications. Without such a system in place at launch the vast majority of the planned DESDynI-R mission data will be under utilized.



**Figure 2. QuakeSim portal showing anonymous disloc version with fault map browser.**

In an end-to-end computational infrastructure users should be able to evaluate data, develop science models, produce improved earthquake forecasts, and respond to disasters in intuitive map-based interfaces. Fault models can be constrained and improved not just by geology, but also by feature identification from InSAR (UAVSAR) and inversions of both GPS and InSAR crustal deformation data [2]. Forecasting is improved by development of better interacting fault models, pattern analysis, and fusion of both seismicity and crustal deformation data. Increasing the accessibility and utility of GPS, InSAR, and geologic data, addresses science challenges such as earthquake forecasting or fluid migration. Intuitive computational infrastructure also can enable new observations by providing tools to conduct simulation experiments and new information products for use in a wide variety of fields ranging from earthquake research to earthquake response. Timely and affordable delivery of information to users in the form of high-level products is necessary for earthquake forecasting and emergency response, but it also necessary for exploiting crustal deformation to enable new discoveries and uses.

There are numerous practical issues to establishing an effective computational infrastructure. Of chief importance is that tools be intuitive and easily accessible. Some QuakeSim tools are public and reside outside of any required login. This mode of operation is often preferred by users as it avoids the need to remember another login and password combination and allows for greater privacy. However, there are also limitations. Chief of these is that project tracking is not possible. The user would be required

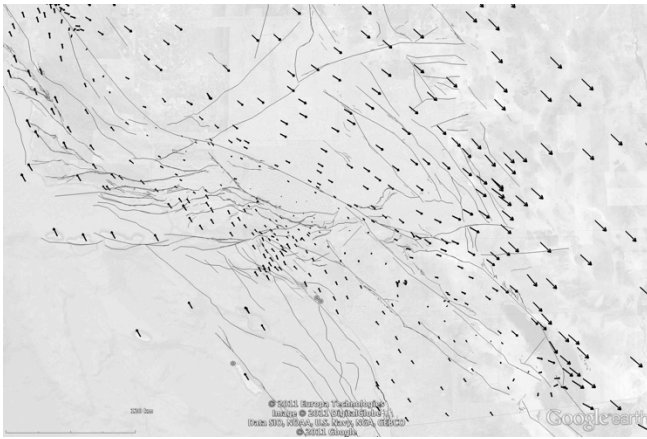
to maintain projects locally, which is reasonably easy with simple input and output files, but becomes rapidly complicated when project components are coupled to various applications at the back end. For example, a user may set up and run a model, which is then coupled to various output format and map views. Linking project components is easier if it is done at the back end within a logged in environment. It also is more efficient in an environment where large data sets are accessed and/or displayed.

Data-intensive computing infrastructure provides a modeling and visualization environment to a broad geophysical community, supporting multiple data types without the need to download large data sets. Access to GPS, InSAR, faults models, and seismicity is just starting to be coordinated: today, large amounts of data must move to the investigator's computer, and integration into models is ad hoc. Modeling interacting-fault simulations largely takes place on local efforts at the research group level, with comparisons taking place largely at infrequent workshops. Web-service based interfaces allow public, independent verification and comparison of simulators and statistical forecast methods, feeding directly into regional hazard models.

Data discovery is also ad hoc and likely to miss important elements of data fusion and cross validation. Ontology-based methods allow immediate discovery of topically or space-time related data. Rather than bringing all data to the user, a system will increasingly need to place substantial processing in cloud computing services close to original or mirrored data archives. Designing around systems from data to investigators will encourage widespread use of enormous data collections such as gathered by radar missions, rather than ad hoc use by a small community of experts. It will be increasingly necessary to couple computing capabilities to the data storage using cloud computing approaches for management of very large data sets.

#### 4. INTERFACES TO EXTERNAL APPLICATIONS

It is not likely, nor is it necessarily desirable, for one monolithic computational infrastructure to be developed for accessing and modeling remotely sensed geodetic data. Centers of expertise are distributed geographically and these centers are providers of components of data, data products, models, or information that lie upstream or downstream from other components. For example, fault modeling connects observational data sets to downstream simulation and forecasting techniques. Ideally these various centers of expertise will develop computational infrastructure that interfaces with upstream or downstream data, data products, and applications. Such a system does not currently exist to support solid earth research, but steps are being taken in that direction as groups realize the need to interface with these upstream and/or downstream components.



**Figure 3. GPS vectors plotted for southern California relative to station CHIL in the San Gabriel Mountains. Fault traces are from UCERF-2 [4].**

NASA's Distributed Active Archive Centers (DAACs, <http://nasaads.eos.nasa.gov/>) process, archive, document, and distribute data from NASA's various missions and field activities. The DAACs serve one or more specific Earth science disciplines and provide data products, meta-data, services, and tools unique to its particular science. Historically analysis was done close to the DAACs. However, as tools and analysis groups become more distributed it is important to ensure that, where necessary, data are mirrored or are otherwise close to the applications to avoid lengthy transfers of large data sets to accomplish tasks.

## 5. VISUALIZATION

Visualization is often necessary for interpreting data or models. Challenges exist both in visualizing complex data as well as in producing visualizations that are properly constrained by data. Visualizations should also be flexible so that the user can view the data or model output in different ways. For example, GPS velocities vectors, when plotted relative to different stations illuminate different features responsible for the deformation (Figure 3). When GPS vectors are plotted relative to the San Gabriel Mountains compression to the west in the Ventura basin becomes clearly apparent. Shear zones on either side of the San Gabriel Mountains are also apparent. Similar plots, but relative to stations in the Mojave Desert highlight the Eastern California Shear Zone.

Movies of deformation can show transient as well as secular deformation. Visualizations are useful when very small changes in the actual observation data are exaggerated in order to be visible to the viewer, and to show a very long timeframe with changes displayed in a compressed time, as long as the exaggeration ratio is chosen properly. Challenges arise with accurately driving the animation from the data. GPS stations are sparsely located as indicated in Figure 3 and as a result interpolations between the stations must be made. Additionally, GPS time series do not all exist for exactly the same time frame, which introduces

meshing complexity or adds artifacts to the visualization from station outages. GPS time series must be properly interpolated both spatially and temporally to provide the most physically accurate animation. InSAR data are also sparse and occur typically for short timespans, but can further guide mapping of crustal deformation. UAVSAR observations in southern California have identified numerous localized zones of shear that could not be identified with the spatial sampling provided by GPS.

## 6. SCIENCE PROBLEMS AS A DRIVER

Basic and applied science tasks illustrate diverse needs for coupling tools and data. Carrying out science problems within the computational environment can help to identify issues and is effective for further development of tools and functionality. Enough tools must be available however to make this appealing to the user. The developed tools can be valuable to end-users as well, however, end-users likely have lower tolerance for tools that don't immediately address their needs and may be less inclined to adopt their usage if they don't work well at the outset. The ideal scenario is to develop the tools with friendly users and then develop documentation and expand the user base as the tools become more functional.

Working with science use cases comes with a different set of challenges. Whereas end users (e.g. for response) need a well-developed set of tools that can be used routinely, science studies often deviate from routine tool usage. Part of the scientific process involves exploring data or models in new ways. As a result keeping up with new tools to satisfy ever changing scientific approaches is challenging. Scientists need toolboxes to develop new approaches more than standardized workflow and the infrastructure and personnel needed to develop toolboxes that allow for flexible analysis of the data is quite extensive. Use cases grouped under three modes: science understanding, forecasting, and response illustrate different user needs and the potential interaction with computational infrastructure and tool developers.

*Science Understanding:* Scientist identifies regions of active crustal deformation from GPS and InSAR/UAVSAR data products. GPS products can be in the form of position time series or station velocities. The scientist scans through the velocity data plotted in vector form on a map in different reference frames to guide thinking as to where active crustal deformation is occurring. The scientist inverts crustal deformation data for fault motions constrained by paleoseismic fault data and then develops simulations based on fault locations and behavior. The scientist may search GPS time series for transient anomalies that indicate previously unknown characteristics of crustal behavior. The possibilities are numerous and the scientist generally wants to explore the data in new ways. Many steps are routine, however, and a friendly scientific user can assess which tasks are carried out frequently enough to warrant new tool development.

*Forecasting:* Scientist identifies active faults from multiple data sources such as GPS, UAVSAR, InSAR, paleoseismic fault data, or seismicity. This is likely to be an outgrowth of the above scientific understanding and exploration. Once techniques are developed pattern analysis is carried out to search for anomalous features in GPS time series and seismicity data. Interacting faults are simulated and statistical analysis of the interactions is conducted. Earthquake probabilities are evaluated for short to decade time scales. Ultimately these probabilities are integrated into the Uniform California Earthquake Rupture Forecast (UCERF). The analysis techniques must be well understood and well defined or standardized in order to incorporate the probabilities into UCERF, which is an official earthquake probability model published by the US Geological Survey [4].

*Response:* When an event occurs deformation can initially be estimated from models that use available seismic information. Initially that information is location, depth, and magnitude of the event. As a result assumptions must be made about the possible mechanism. Where fault data are available the likely mechanism can be constrained to known faults. In time an earthquake mechanism is produced, which provides two orthogonal geometries of slip. The deformation estimates can be used to estimate the envelope of maximum displacement, and hence most likely region of damage. This envelope can also be used to guide acquisition of UAVSAR and GPS data for emergency and science response. Possible locations of future aftershocks can be assessed, as the fault models are refined. The damage zone from the event can be defined as a polygon and formatted or ingestion into loss estimation tools. Damage and potential aftershock assessments can be refined as new information becomes available. The products can be made available to emergency responders. Such products must be easily accessible and intuitively interpretable by responders. It is highly important that any tools be bug free and support routine uses.

## 7. CONCLUSIONS

Increasing data volumes, complexity of data processing algorithms, and more comprehensive models continually drive a need for more and more compute power. Additionally, as large data sets are accessed it is necessary to either keep the data close to the models or have extremely broadband connections between the data sources and computers carrying out the models.

The largest jobs are currently run on supercomputers, which reside at high performance computing facilities. These resources are often oversubscribed and users' models can spend a long time in a queue before the job runs and investment needs to be made in more high performance computers. QuakeSim is architected so that it can scale up by the many orders of magnitude needed when new satellites or other instruments result in much larger data sets. Iterative MapReduce [5] that interoperates between HPC

and cloud environments can be deployed to handle the much larger datasets and model runs. We have developed the OpenQuake concept [9] which is an open interface portal aimed especially at real time needs of science and crisis response with both commercial and research data sources and users. Note that most data analysis applications are well suited to cloud environments and can exploit the elasticity and cost-performance advantages of clouds. The real-time needs of major events can require especially elastic computing on demand hosted by OpenQuake.

Visualization tools are increasingly necessary for understanding data and models. Users are frequently hampered by tools that rely on licensed products or not run in the same environment in which the data or models are stored or processed. This results in the need to move large volumes of information and often requires an additional reprocessing or reformatting step before visualization can take place. Current open source tools for this are not yet mature. An investment in open source visualization tools will result in much greater scientific efficiency. Both data management and simulation tools would benefit from a redesign of the underlying computing infrastructure.

## ACKNOWLEDGEMENTS

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, Indiana University, University of Southern California, and University of California's Davis and Irvine campuses under contract with NASA. The work was sponsored by NASA's Advanced Information Technologies Program. We thank UNAVCO, SOPAC, ASF, and University of Nevada, Reno for ongoing collaborations.

## REFERENCES

- [1] Donnellan, A.; Rosen, P.; Graf, J.; Loverro, A.; Freeman, A.; Treuhaft, R.; Oberto, R.; Simard, M.; Rignot, E.; Kwok, R.; Xiaoqing Pi; Blair, J.B.; Abdalati, W.; Ranson, J.; Zebker, H.; Hager, B.; Shugart, H.; Fahnestock, M.; Dubayah, R.; , "Deformation, Ecosystem Structure, and Dynamics of Ice (DESDynI)," *Aerospace Conference, 2008 IEEE* , vol., no., pp.1-13, 1-8 March 2008 doi: 10.1109/AERO.2008.4526249
- [2] Donnellan, A., Parker, J.W. and Peltzer, G., Combined GPS and InSAR models of postseismic deformation from the Northridge earthquake, *Pure Appl. Geophys.*, **159**, 2261-2270 (2002)
- [3] Wei, M, D Sandwell and B Smith-Konter, Optimal combination of InSAR and GPS for measuring interseismic crustal deformation, *Adv. Space Res.*, **46**, 236-249 (2010).

- [4] Field, E.H., T.E. Dawson, K.R. Felzer, A.D. Frankel, V. Gupta, T.H. Jordan, T. Parsons, M.D. Petersen, R.S. Stein, R.J. Weldon II, and C.J. Wills, The Uniform California Earthquake Rupture Forecast, Version 2 (UCERF 2), By 2007 Working Group on California Earthquake Probabilities, USGS Open File Report 2007-1437, CGS Special Report 203, SCEC Contribution #1138, 2008.
- [5] SALSA Group. Iterative MapReduce. 2010 [accessed 2010 November 7]; Twister Home Page Available from: <http://www.iterativemapreduce.org/>.
- [6] J.Ekanayake, H.Li, B.Zhang, T.Gunarathne, S.Bae, J.Qiu, and G.Fox, Twister: A Runtime for iterative MapReduce, in Proceedings of the First International Workshop on MapReduce and its Applications of ACM HPDC 2010 conference June 20-25, 2010. 2010, ACM. Chicago, Illinois. <http://grids.ucs.indiana.edu/ptliupages/publications/hpd-c-camera-ready-submission.pdf>.
- [7] Bingjing Zhang, Yang Ruan, Tak-Lon Wu, Judy Qiu, Adam Hughes, and Geoffrey Fox, Applying Twister to Scientific Applications, in CloudCom 2010. November 30-December 3, 2010. IUPUI Conference Center Indianapolis. <http://grids.ucs.indiana.edu/ptliupages/publications/PID1510523.pdf>.
- [8] Thilina Gunarathne, Judy Qiu, and Geoffrey Fox, Iterative MapReduce for Azure Cloud, in CCA11 Cloud Computing and Its Applications. April 12-13, 2011. Argonne National Laboratory, Chicago, ILL. <http://grids.ucs.indiana.edu/ptliupages/publications/cca-v8.pdf>.
- [9] Geoffrey Fox, OpenQuake Infomall, APEC Cooperation for Earthquake Simulation (ACES) Workshop on Advances in Simulation of Multihazards Maui, Hawaii, May 1-5, 2011 on May 4 2011 [http://grids.ucs.indiana.edu/ptliupages/presentations/ACES\\_May4-11.pptx](http://grids.ucs.indiana.edu/ptliupages/presentations/ACES_May4-11.pptx)

## BIOGRAPHIES

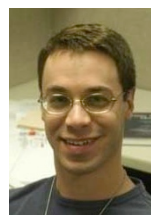


**Andrea Donnellan** has been a geophysicist at NASA's Jet Propulsion Laboratory since 1993, and a research professor at the University of Southern California since 1998. Donnellan studies earthquakes and crustal deformation by integrating satellite technology with high performance computer models. She is Principal Investigator of NASA's QuakeSim project, as well as supercomputing, earthquake modeling, and UAVSAR projects. Donnellan has been Deputy Manager of the JPL's Science Division, Pre-Project Scientist of a mission to study natural hazards, ice sheets, and ecosystems, and NASA's

Applied Sciences Program Area Co-Lead for Natural Disasters. She has conducted field studies in California, in Antarctica, on the Altiplano of Bolivia, in Mongolia, and on Variegated Glacier in Alaska. Donnellan received a bachelor's degree in geology with a mathematics minor from the Ohio State University in 1986. She received her master's and Ph.D. in geophysics from Caltech's Seismological Laboratory in 1988 and 1991 respectively and held a National Research Council postdoctoral fellowship at NASA's Goddard Space Flight Center from 1991-1993. Donnellan received an M.S. in Computer Science from the University of Southern California in 2003.



**Jay Parker** joined the Satellite Geodesy and Geodynamics Systems Group at JPL in 1996, and has been part of the JPL technical staff since 1989. He completed both a master's and PhD in Electrical Engineering from the University of Illinois (Urbana-Champaign), and graduated with a Bachelors of Science from the California Institute of Technology in 1981. His professional interests lie in applications of fast and accurate numerical models to geophysical remote sensing. Past modeling projects include vortex formation in the ionospheric D region, parallel supercomputer modeling of radar scattering and antenna power patterns, and high-fidelity modeling of clear-air infrared spectra for determining climate change and pollution sources. He is currently working on methods to invert SCIGN GPS data to determine earthquake and after-slip fault movements, finite element models of earthquake cycles, and new methods for GPS data processing on supercomputers. Jay has been inducted into Tau Beta Pi, and received a JPL Technology and Applications Programs Group Achievement Award. He is a member of the American Geophysical Union, and the IEEE Antennas and Propagation Society. Outside of work Jay enjoys exploring the local mountains, reading 19th century literature, and collecting folk music albums.



**Robert Granat** is currently Group Supervisor of the Machine Learning and Instrument Autonomy group at JPL, and has been a member of technical staff there since 1996. He received his M.S. and Ph.D. degrees in Electrical Engineering from the University of California, Los Angeles, and his B.S. from the California Institute of Technology. Since 1999, he has been working on QuakeSim and related projects to perform statistical machine learning based health monitoring, signal classification, and anomaly detection on seismic and GPS sensor networks. His other research interests include autonomous robotic navigation, scalable scientific computing, and radiation fault tolerant algorithms for spaceborne computation.



**Eric M. De Jong** is Chief Scientist for JPL's Instrument Software and Science Data Systems Section; and Research Director for JPL's: Visualization and Image Processing (VIP) Center, Image Processing Laboratory (IPL), Digital Image Animation Laboratory (DIAL), and Cartographic Analysis Laboratory (CAL).

Eric is a Planetary Scientist in NASA's Jet Propulsion Laboratory Science Division and a Visiting Associate in Planetary Science at Caltech. For the last three decades his research has focused on the scientific visualization of the Earth; Sun; planetary surfaces, atmospheres, magnetospheres; and the evolution and dynamics of stars, galaxies and planetary systems. As the Principal Investigator for NASA's Space and Earth Science Visualization (SSV) Project he leads a team of scientists and technologists responsible for developing new science visualization products, infrastructure, technology, tools and services. De Jong and his team create movies, images, mosaics, maps and models from NASA Space & Earth Science remotely sensed data. These products highlight discoveries, science results, mission plans and operations. He received his B.S. in Plasma Physics from the Massachusetts Institute of Technology in 1967; M.S. in Plasma Physics from Stanford University in 1967, Ph.D. in Interdisciplinary Science from The University of California, Santa Barbara in 1982, and served as a Post-Doc in Planetary Science at Caltech in Pasadena Ca. in 1989.



**Shigeru Suzuki** is a member of the technical staff at NASA's Jet Propulsion Laboratory. For the last two decades, he has been working on visualizing the science data for NASA's Space and Earth Science Visualization (SSV) Project, where he has developed new science visualization products, infrastructure, technology, tools

and services. He received a BsEE from Tokyo University of Agriculture and Technology (1979). Before joining JPL in 1991, he worked for the Japan Broadcasting Corporation (NHK) as a video engineer.



**Geoffrey Fox** received a Ph.D. in Theoretical Physics from Cambridge University and is now professor of Informatics and Computing, and Physics at Indiana University where he is director of the Digital Science Center and Associate Dean for Research and Graduate Studies at

the School of Informatics and Computing. He previously held positions at Caltech, Syracuse University and Florida State University. He has supervised the PhD of 62 students and published over 600 papers in physics and computer science. He currently works in applying computer science to Bioinformatics, Defense, Earthquake and Ice-sheet Science, Particle Physics and Chemical Informatics. He is principal investigator of FutureGrid – a facility to enable development of new approaches to computing. He is

involved in several projects to enhance the capabilities of Minority Serving Institutions.



**Marlon Pierce** is the Assistant Director for the Science Gateways Group in Research Technologies Applications at Indiana University. Pierce received his Ph.D. Florida State University (Physics) in 1998 in computational condensed matter physics. His current research and development work focuses on

computational sciences with an emphasis on Grid computing and computational Web portals. Prior to forming the Science Gateway Group, Pierce served as assistant director for the Community Grids Laboratory at Indiana University's Pervasive Technologies Institute. Pierce supervises the research activities of software engineering staff and Ph.D. students, and serves as principal investigator on multiple federally-funded research projects. Pierce leads research efforts in the following areas: the application of service-oriented architectures and real-time streaming techniques to geographical information systems and sensor networks; the development of open source science Web portal software for accessing Grid computing and data resources; and Grid-based distributed computing applications in computational chemistry and material science, chemical informatics, and geophysics.



**John Rundle** is an Interdisciplinary Professor of Physics, Civil Engineering and Geology at University of California, Davis. His research is focused on understanding the dynamics of earthquakes through numerical simulations; pattern analysis of complex systems; dynamics of driven nonlinear

Earth systems; and adaptation in general complex systems. Computational science and engineering is an emerging method of discovery in science and engineering that is distinct from, and complementary to, the two more traditional methods of experiment/observation and theory. The emphasis in this method is upon using the computer as a numerical laboratory to perform computational simulations to gain insight into the behavior of complex dynamical systems, to visualize complex and voluminous data sets, to perform data mining to discover hidden information within large data sets, and to assimilate data into computational simulations. Professor Rundle is a Fellow of the American Geophysical Union.

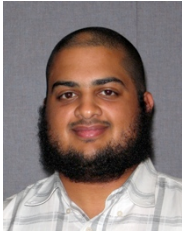


**Dennis McLeod** is currently Professor of Computer Science at the University of Southern California, and Director of the Semantic Information Research Laboratory. He received his Ph.D., M.S., and B.S. degrees in Computer Science and Electrical Engineering from MIT. Dr.

McLeod has published widely in the areas of data and knowledge base systems, federated databases, database



models and design, ontologies, knowledge discovery, scientific data management, information trust and privacy, and multimedia information management. His current research focuses on: structured domain ontologies; semantic web; database semantic heterogeneity resolution and inter-database correlation; personalized information management and customization; information management environments for Earth, marine, and climate science; the architecture of data centers providing massive storage via virtualization and data clouds; social networking information management and information trust; and service-based information access and delivery frameworks.



**Rami Al-Ghanmi** is a Ph.D. student in the Computer Science Department at the University of Southern California. He received his MS in Computer Science from the USC (2006) and BS in Computer Engineering from King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia (2002). He is currently working with Professor Dennis McLeod at the Semantic Information Research Group at USC. His research interests are semantic web services and ontology-based federation of scientific data.



**Lisa Grant Ludwig** is an Associate Professor in Public Health at the University of California, Irvine. She was Associate Director of the California Institute for Hazards Research. Lisa earned a Ph.D. from Caltech in 1993 in Geology with Geophysics, M.S. degrees from Caltech in 1990 in Geology and in 1989 in Environmental Engineering and Science, and a B.S. from Stanford in 1985 in Applied Environmental Earth Sciences. Lisa's research interests include natural hazards, paleoseismology, active faults, San Andreas fault, southern California faults, San Joaquin Hills, seismic hazard, environmental health and geology. Her research group addresses natural hazards and disasters from a geologic perspective, with emphasis on earthquakes. Earthquakes are a major threat to public health globally and locally in California. The group focuses on defining the potential for large earthquakes, and working collaboratively on developing forecasts, hazard models and effective responses. Results of the work are applied for disaster preparedness planning, structural design, land-use planning, seismic risk assessment and public education about earthquake hazard.