Multi-Hazard Data Science:

A Workshop

Sponsored by

University of California, Davis

Under the Auspices of the Association of Pacific Rim Universities

June 26-28 2019, Davis, California

APRU
Multi-Hazards

Contents

Forward

A primary focus of the workshop was to understand how the methods of data and decision science in the commercial and business worlds can be used to improve societal resilience in the face of increasing threats from natural disasters. In addressing this issue, the workshop focused on the seven targets outlined in the Sendai Framework[1] for Risk Reduction, particularly Target g:

Substantially increase the availability of and access to multi-hazard early warning systems and disaster risk information and assessments to the people by 2030

A second focus was the Asia–Pacific Economic Cooperation goal of enhancing expertise in data competency, through an emphasis on data science and analytics[2].

[1]https://www.unisdr.org/we/coordinate/sendai-framework
[2]https://www.apec.org/Press/Infographics/10-Recommended-APEC-Data-Science-and-Analytics

Goal

The overarching goal of the workshop was to better understand how data and decision science can impact each of the four stages in the evolution of natural disasters:

- **Anticipation**
- **Mitigation**
- **Response**
- **Recovery.**

Each of these four areas was the subject of interdisciplinary sessions at the workshop.

Local Organizing Committee:

John Rundle
jbrundle@ucdavis.edu

Ken Verosub
klverosub@ucdavis.edu

Lorraine Hwang
ljhwang@ucdavis.edu

Duncan Temple-Lang
dtemplelang@ucdavis.edu

Mark Yoder
mark.yoder@gmail.com

Introduction

In spring 2019, the American Academy of Arts and Sciences released a report entitled: "*Science During Crisis: Best Practices, Research Needs, and Policy Priorities*[3]". Authored by R. Colwell and G. Machlis, the report detailed a number of critical areas for improved response and analysis of great natural disasters. One of these areas was the use of "Big Data to Support Science During Crisis". Recognizing that greatly improved data is necessary to support all facets of disaster research, the Organizing Committee placed the focus of the workshop on improvements in data and decision science. This emphasis is congruent with the growing interest within the larger Asia Pacific Economic Cooperation (APEC) and Association of Pacific Rim Universities (APRU) for the development of data science and data analytics expertise among its member groups.

As a result, the workshop was concerned with addressing the issues posed in the above report:

- *How can data science best address the four phases of anticipation, mitigation, response and recovery?*
- *What are the data products needed, what data is needed to create them, and what time frames are required to produce these products?*
- *Can better data science stimulate and encourage people to appreciate their risk?*
- *What comprises science and data science during crisis?*
- *How do we establish baseline information?*
- *How do we deal with cascading consequences to improve response and rebuilding?*
- *How do we address divergent scientific opinions, data and results during crises especially, but also in the other phases of disasters?*
- *How do we best communicate during crises?*
- *What is the best way to make science-based decisions before, during, and after crises?*
- *How can we use big data to support science during crises?*

[3] R.R Colwell and G.E. Machlis, Science During Crisis: Best Practices, Research Needs, and Policy Priorities, Amer. Acad. Arts Sci., 2019. https://www.amacad.org/events/science-during-crisis

Workshop Organization

The workshop was organized as a series of panel discussions on the selected topics of Anticipation, Mitigation, Response and Recovery. Panel members were asked to limit their prepared remarks and slides so that maximum time could be reserved for extended discussions, in contrast with the more usual workshop methods of only a series of prepared talks followed by a few questions. Thus the primary mode of interaction was open discussion on each topic.

To set the stage, a keynote address was presented by Gary Machlis of Clemson University, one of the principal authors of the AAAS report on Science During Crisis. His presentation was titled "*Science During Crisis, and its Distinctive Demands on Data Science*". This was followed by Christina Schönleber, Director Policy & Programs providing an overview of APRU and the Associations main activities, and Professor Takako Izumi from Tohoku University introducing the APRU Multi-Hazards Hub and its key program activities based at Tohoku University in Sendai, Japan. Professor Duncan Temple-Lang then described the Data Science Initiative at UC Davis, followed by Professor Bruce Malamud of Kings College, London, presenting a discussion of "*Multihazards and Hazard Interrelationships in Developing Countries*". Thereafter, the general plan was to empanel the participants in discussions of domain science and data science on the relevant topics. Moderators directed the flow of discussions, and Student Reporters recorded the particulars of the discussions. What follows in the description of panel discussions here is drawn from the notes of the Student Reporters (A. Giguere, R. Fildes, B. Chap, D. Grzan, C. Saylor, and R. Ortez).

Gary Machlis

Bruce Malamud

Panel on Anticipating Multihazards - Domain Science

Moderator: K. Verosub

Student Reporter: C. Saylor

Panelists:
Masakazu Hashimoto, Lev Kavvas, Tony Song, Attila Komjathy, Panagiotis Vergados, Mike Oskin, Bruce Malamud, John Rundle

Suggested Topics Included:
Earthquakes, Floods, Tsunamis, Wildfires, Landslides

Summary

- The single most discussed topic was the poor communication between scientists and those in disaster response/mitigation
- Suggestions to improve communication included planning meetings between scientists and disaster response
- Another suggestion was to recommend use of a clearinghouse that acts as liaison between the two communities

Discussion Points

- Agencies do not make science based decisions, they make at most science informed decisions, where science is only one component
- They look to see if scientific information will hold up in court if a decision is made based on that information
- Scientists and those in disaster research must become proficient and learn about the political process
- Agencies are always calculating what is the "precedent danger" of following one policy or another?
- Agencies prefer to use procedures that don't have legal consequences

Best Practices

- Reach out to agencies for meetings to socialize information ahead of time, this must happen BEFORE a disaster
- Determine how to establish a sort of symbiotic relationship that supports effectiveness of both parties without inhibiting either
- Learn from a position of mutual respect
- Different products may need to package the same information for different groups of people
    - Determine in advance how can information best be packaged, simplified, presented

    - Realize that many groups would find use in the same information at different points in the hazard prevention/recovery process
- Establish a training period/retreat process where politicians/decision-makers can meet with scientists to discuss what is being worked on that could be useful to the other party

Panel on Anticipating Multihazards - Data Science

Moderator: Lorraine Hwang

Student Reporter: Bottreypich Chap

Panelists:
Mark Yoder, Jay Parker, John Wilson

Suggested Topics Included:
Multihazard Data Science, Simulations, Remote Sensing, Image analysis

Summary

- A main theme included data science as data informed discovery, leading to a better understanding of the goals for data collection
- Prioritize the discovery of datasets that are available, measurable, and pertinent
- Focus now and increasingly in the future on scientific **machine learning** and **decision science**

Discussion Points

- A major focus (but not the only focus) of data science is forecasting, nowcasting, and anticipation of major disasters
- Uncertainty in both the data and the models must be shown visually and clearly to both agencies and stakeholders to minimize misunderstandings
- Reliability of current models should be established for a given phenomenon
- Decision makers need to understand the degree of certainty in order to make policy
- How should resources best be utilized? Budgets are always inadequate so prioritizing resources is critical
- Scientists, stakeholders, and decision makers should all speak a common language relating to the communication of uncertainty

Best Practices

- Uncertainty should be communicated:
    - Using language that policymakers and stakeholders are familiar with
    - Utilizing visuals that are easily interpreted without the data creator
    - Showing before and after photos for relative comparisons
    - Along with recommendations for focal points for the following recovery process.
- A significant need exists to correct misconceptions about natural hazards, for example the idea that aftershocks are weaker and less destructive than the initial earthquake.
- New directions in disaster forecasting should be emphasized, including:
    - Time dependent confidence bands

- Recognition of different time scales for different hazards such as earthquakes, tsunamis and tornadoes.
- Analysis that should not be purely statistical, but rather carried out in conjunction with growing observations to better fit the data.

Panel on Mitigating Multihazards

Moderator: Mark Yoder

Student Reporter: Ronaldo Ortiz

Panelists:
Sashi Kunnath, Tom Beamish, Eric Heien, Ai-Lei Sun, Erik Porse, Richard Armstrong

Suggested Topics Included:
Engineering for Resilience, Nature of Risk, Engineering Data Science

Summary

- Discussion focused on the possibility of using various satellite imaging techniques to assess extent of damage in affected areas.
- Processes can now be automated with requests being met within 24 hours (now)
- Questions surround censorship of certain customer bases, and led to the larger question of how much privacy should be forfeited to improve safety

Discussion Points

- There might be a trade-off between privacy and improvements in safety
- But even without compromise, such tools could be used for mass infrastructure assessment.
- As an example, rapid satellite imaging could be employed to quickly identify safe evacuation routes and particular transportation failures that could hamper emergency response efforts.
- Systems used to anticipate earthquake damage were discussed.
- In particular, widely used systems like ShakeCast should be refined in order to better anticipate damages.
- Dams seem to have particular vulnerabilities from earthquakes both in how damages are monitored and lack sufficient categorization of damages done.
- The role of utility companies and the procedures they might use in order to respond to hazards were also discussed, no conclusions reached.

Best Practices

- Satellite imaging is now routinely used for event identification, including crop yields, ship location/ illegal fishing breaking embargos
- Imaging provides insight to different customers: governments, industry, insurance companies
- Efficient algorithms should be developed to find and identify objects such as buildings, railways, land use, ships, airplanes, to include methods such as:
    - Calibration on training data to train the neural net

    - Use of optical and radar image data sets together as a means of data validation
- Other imaging products could include:
    - Flood impacts such as from Hurricane Harvey
    - Extent of wildfires such as in Northern California:
    - Flood Monitoring in locations such as Japan, to include answers to questions such as: Where is the flood and how deep is the water?
- Algorithm Development:
    - Human "marking" can be used to train the neural nets
    - It is difficult to visually identify Japanese buildings and population densities, perhaps neural nets can do better.
    - Uses optical and radar image data sets.
    - Needs ground truth as well
- With respect to earthquakes and damage assessment:
    - Need to work with the affected agencies such as Caltrans
    - Use highly nonlinear Finite Element models to generate estimates of potential loss
    - Develop fragility models from data and simulations
- Earthquake challenges include:
    - Defining an intensity measure
    - Determining intensity of shaking at a site
    - Using spectral quantities to characterize damages
    - Loss modeling
    - Estimating non-structural damage, downtime
    - Classifying structures:
    - Analyzing data in groups: mid-rise, high-rise, type, material, bridges
    - Determining construction era, codes
    - Risk assessment from single structure to regional or city scale.
    - Propagating uncertainties
- Dam structure challenges include:
    - Understanding impact of earthquakes on dams: 10% dams are instrumented,, and data are publicly available
    - Hydrological information is typically separate from other structural data
    - There is the possibility of using machine learning for hydro data to learn more about dam behavior
    - ShakeCast is used by regulators and dam owners.
    - Multihazards: A dam hazard is analyzed separately, but should be analyzed in context with other types of hazards
    - Defining what constitutes a dam failure? Uncontrolled release?
    - Typically there is no loss modeling so quantifying damage is needef
    - For mitigation, failure mode analysis is needed together with an emergency action plan

Panel on Responding to Multihazards

Moderator: John Rundle

Student Reporter: Rebecca Fildes

Panelists:
Ken Verosub, Mike Sjoblom, Nick Anderson, Tom Nesbitt, Sebastian Schwindt

Suggested Topics Included:
Seismic Damage Assessment, Disaster Informatics, Data Science Analytics

Summary

- Data aggregation and integration for consumer health data is a critical need
- Rapid services are needed for damage assessment, based on automated sensor webs and internet communication
- There is a need for continuing disaster drills

Discussion Points

- Connections should be built with government agencies before hazards occur in order to more quickly gain support for rapid science response after a disaster occurs
- FEMA: Disaster Recovery Reform Act (2018) covers many of the relevant aspects
- There is a need for involvement in multi-hazard drills such as National Guard puts on "Golden Guardian" and other state level exercises:
  - A general scenario is defined and exercise participants decide how to participate.
  - Insight is needed into what decision makers are asking, what kind of data they need, how it is best presented and the workflow on the military side of disaster response.
- How can different data types be combined effectively for decision makers?
- "The disaster is not just the day of the event":
  - It is not just first aid immediately after a hazard (or multiple hazards) strikes an area, but continuous help after the initial event needs to be thought about such as where to set up temporary pharmacies
  - Cartographers can help with effective and efficient presentation of data. Maps have to be scalable and readable. They need to be easy to read with clearly illustrated and defined data for easy distribution and interpretation during hazardous scenarios
  - Modeling and simulations of multi-hazard events and interactions: machines can run exercises all the time, to keep things from getting "stale" , simulations should be run frequently, and updated as new information, and more events/examples occur

Best Practices

- Aggregating health data can get remove the sensitivity/privacy issues, but restrictions as to where and how the data is available affects the usefulness of the data
- Linking people to health conditions during disasters could lead to checking in on people during disasters, and provide information for self-management if people are inaccessible
- A major issue in many disasters is simply determining the amount and extent of damage rapidly so that appropriate response planning and aid delivery can be carried out rapidly

Panel on Recovery from Multihazards

Moderator: Ken Verosub

Student Reporter: David Grzan

Panelists:
Takako Izumi, Anne Rosinski, Phil Beilin

Suggested Topics Included:
Seismic Damage Assessment, Disaster Informatics, Data Science Analytics

Summary

- A major focus was on lesson learned from the 2011 Tohoku earthquake and tsunami, including the concept of "Multiple Defense", i.e., creating a tiered defense system
- To aid in recovery, funding must be obtained by strategically approaching state and federal funding agencies
- Data must be formatted in ways that allow decision makers and the public to easily absorb it.

Discussion Points

- Multiple defense against tsunamis in Japan following the March 11, 2011 earthquake consists of building embankments, elevated roads, structures that allow for "escape in place", together with moving residential areas and schools to higher elevations
- As an example, the town of Okatsu constructed a 9.7 meter seawall, and built structures into the local hillsides
- The latter strategy has also been adopted by the town of Onegawa, which experienced a ~20 meter tsunami in 2011
- Teachers should be better prepared to lead their classes in emergency procedures, which was not the case in Japan in 2011 in the Okawa elementary school - most of the children died because the teachers thought the water would not reach the school, and so no evacuation was carried out
- Mega-disasters are not becoming more common because the events are getting larger, rather it is because more people are moving into risk-prone areas
- Funding to aid in recovery should be approached strategically in a rule-based manner, rather than tactically, and it is important to demonstrate in advance that useful actions can emerge
- State and local agencies interact with data differently than scientists, and formatting data for different groups should be considered

- For scientists to talk to decision makers, communication of critical information needs to be based on why they should care, and how they should use the information
- There is no standard mechanism for optimal communication, although some fields are better than others (climate change methods were offered as a example of how other fields might approach this problem)
- Face-to-face communication is particularly important
- Methods to optimize presentations include clever use of colors, simplicity of presentation, and emphasis on the visual

Best Practices

- Improvement of teacher education and disaster preparedness should be emphasized
- Training in optimal communication practices during times of disasters is critical
- Reduction and presentation of critical data to non-specialists should be given much more attention than it is at present

Challenges and Recommendations

(Panel on Where do We Go from Here?)

Moderator: Mark Yoder

Student Reporter: Alexis Giguere

Panelists:
Takako Izumi, Elizabeth Langridge, Phil Beilin, Jeff Onstedd, Gary Machlis, Bruce Malamud

Summary

- Discussions centered around integration of data science practices with policy-making

Discussion Points

- *Communication between science and policy*. This is a problem at every level of analysis, from job postings to disaster data interfaces.
- While there are programs designed to facilitate collaboration between science and policy, there is often *no follow through*. People meet and discuss, but they then return to their own fields after concluding the discussion
- *Institutional structures are not always conducive to interdisciplinary work.* Stratification of government agencies can work against data science integration goals.
- *Hazard and vulnerability are two distinct concepts that are often used interchangeably*. This is an issue, in the sense that it tends to paper over the dynamical nature of disasters. Hazard, vulnerability, and exposure are often not treated separately, which can lead to a loss of resilience in the system.
- *Social justice aspects can be forgotten in the chaos of the aftermath of a disaster*. Prioritization is implemented according to various metrics, and when the metrics involve a social justice component, they can be difficult to quantify. Can data science provide a means of quantifying social justice needs so that planners and first responders do their job more efficiently and more fairly at the same time?

Best Practices

- *Marshaling the power of private institutions.* There are many examples of private corporations (utilities, insurance companies) whose goals intersect with the goals of the public agencies charged with hazard management. They often have fairly well-developed data pipelines and streamlined decision-making processes that could be useful for data science integration.

- *Mutual respect between the science and policy communities.* It's a misconception that social scientists don't care about data. Everybody needs the data coming out of data science analyses, but the level of analysis can be vastly different. Even if the analysis is done similarly, the language used to frame the policy debate can be a source of friction. Policy is political and politics is cultural, so cultural differences can be barriers to effective data science/policy integration. This means that there is a large premium on 'cross-cultural interpreters'.
- *Use already existing infrastructures more effectively*. A point made repeatedly throughout the session is that there are many existing structures encouraging interdisciplinary collaborations. Institutions need to make themselves known; so do the specific individuals within the institutions who act as gateways.
- *Prioritize effective data dissemination.* Peer-reviewed papers are nice, but there are more ways to effectively conduct knowledge transfer. Blogs, vlogs, informal meetings, and non-academic articles can be used to bridge the link between policy makers and data scientists. This non-traditional way of 'publishing' demands that we pay closer attention to the way we present data.
- *Innovate*. Data science can thrust itself to the forefront of policy discussions if it can show itself to be relevant. Increasingly more powerful techniques (like innovations in natural language processing) have the potential to 'humanize' the data sets and therefore make them more approachable, regardless of what's under the hood. At the same time, the amount of available data is exploding, so data reduction techniques are growing more and more important.

Appendices

I. Program

June 26 (Wednesday) Registration 3:00-5:00 pm

June 27 (Thursday)

9:00 Welcome, Introductions, Goals of the Workshop (John Rundle, Ken Verosub)

9:10 Welcome from UCD Global Affairs (Joanna Regulska, Vice Provost)

9:15 Introduction to the Association of Pacific Rim Universities (APRU)
(Christina Schönleber, Director, Policy & Programs)

9:30 for Introduction to the APRU Multihazards Hub and International Research Institute Disaster Science (IRIDeS) at Tohoku University, Sendai, Japan
(Takako Izumi, Associate Director)

9:45 Data Science (Duncan Temple-Lang, UCD Data Science Initiative)

10:00 Keynote: "Science during crisis, and its distinctive demands on data science" (Gary Machlis, Clemson University)

10:45 Break

11:00 "Multihazards and hazard interrelationships in developing countries"
(Bruce Malamud, Kings College)

11:15 Panel: Anticipating Multihazards - Domain Science
(Moderator: Ken Verosub, Student Reporter, Cameron Saylor)

- Panelists: Masakazu Hashimoto, Lev Kavvas, Tony Song, Attila Komjathy, Panagiotis Vergados, Mike Oskin, Bruce Malamud, John Rundle
- Suggested Topics: Earthquakes, Floods, Tsunamis, Wildfires, Landslides

12:30 Lunch

1:30 Panel: Anticipating Multihazards - Data Science
(Moderator: Lorraine Hwang, Student Reporter, David Grzan)

- Panelists: Mark Yoder, Jay Parker, John Wilson, James Holliday
- Suggested Topics: Multihazard Data Science, Simulations, Remote Sensing, Image analysis

2:45 Panel: Mitigating Multihazards
(Moderator: Mark Yoder, Student Reporter, Ronaldo Ortiz)

- Panelists: Sashi Kunnath, Tom Beamish, Eric Heien, Ai-Lei Sun, Erik Porse, Richard Armstrong
- Suggested Topics: Engineering for Resilience, Nature of Risk, Engineering Data Science

4:00 Break

4:15 Open Discussion

5:00 Reception

June 28 (Friday)

9:00: Panel: Responding to disasters

(Moderator: Takako Izumi, Student Reporter, Becky Fildes)

- Panelists: Ken Verosub, Mike Sjoblom, Nick Anderson, Tom Nesbitt, Sebastian Schwindt, James Wollbrinck
- Suggested Topics: Seismic Damage Assessment, Disaster Informatics, Data Science Analytics

10:30 Break

10:45 Panel: Recovery from disasters

(Moderator: Ken Verosub, Student Reporter, Cassey Chap)

- Panelists: Takako Izumi, Shabir Kabirzad, Anne Rosinski, Phil Beilin
- Suggested Topics: Mega-Tsunamis, City Resilience, Local and National Recovery Programs

12:00 Lunch

1:00 Student Reports on Panels

1:30 Panel: From the local to the international - Where do we go from here? (Challenges and Recommendations)

(Moderator: Mark Yoder, Student Reporter, Alexis Giguere)

- Panelists: Takako Izumi, Joanna Regulska, Phil Beilin, Jeff Onstedd, Gary Machlis, Bruce Malamud
- Suggested Topics: Open

2:30 Break

2:45 Panel: Summary Discussion

(John Rundle, Ken Verosub, Lorraine Hwang, Mark Yoder, Duncan Temple-Lang)

5:00 Closing and End

II. Attendees (Participants)

First Name	Last Name	Affiliation
Nick	Anderson	UC Davis
Richard	Armstrong	CSU, Sacramento
Thomas D.	Beamish	UC Davis
Phil	Beilin	CA Earthquake Clearinghouse IT lead
Magali	Billen	UC Davis
Bottreypich	Chap	UC Davis
Ali	Ercan	UC Davis
Rebecca	Fildes	UC Davis
David	Grzan	UC Davis
Masakazu	Hashimoto	IRIDeS, Tohoku University
Eric	Heien	Orbital Insight
Lorraine	Hwang	UC Davis
Takako	Izumi	IRIDeS, Tohoku University
Levent	Kavvas	Dept. of Civil&Env. Engg., UC Davis
Attila	Komjathy	NASA JPL
Sashi	Kunnath	UC Davis
Elizabeth	Langridge	Global Affairs
Gary	Machlis	clemson university
Bruce	Malamud	Kings College London
Jeffrey	Onsted	California Department of Conservation
Ronaldo	Ortez	UC Davis
Michael	Oskin	University of California, Davis
Jay	Parker	Jet Propulsion Laboratory/California Inst. Tech.
Erik	Porse	Sacramento State ^ UCLA
Joanna	Regulska	Global Affairs, UC Davis
Anne	Rosinski	FEMA Region IX
Efraim	Roxas	Univ. Philippines Los Banos / Florida State
John	Rundle	university of California Davis
Cameron	Saylor	UC Davis
Christina	Schönleber	APRU
Sebastian	Schwindt	University of California, Davis
Abdul	Sesay	Engineer
Tony	Song	NASA/ JPL
Ai-Lei	Sun	Orbital Insight
Panagiotis	Vergados	JPL/Caltech
Kenneth	Verosub	UC Davis
John	Wilson	UC Davis
Mark	Yoder	UC Davis