SUMMARY OF THE NSF CRYOSLIDERISK WORKSHOP, PENN STATE, MAY 12-13, 2022

ABBREVIATED WORKSHOP AGENDA

MAY 11: An informal icebreaker was held at the Happy Valley Brewing Company.

MAY 12 - DAY 1: Machine learning day, introduction of mass movement hazards in cryosphere

- Welcome, overview, anticipated outcomes and products (*Tong Qiu, Civil and Environmental Engineering, Penn State*)
- Overview of AI/ML for civil engineering and geosciences (*Chaopeng Shen, Civil and Environmental Engineering, Penn State*)
- Mass movements in warming permafrost slopes (*Stephan Gruber (remote participation), Geography and Environmental Studies, Carleton University*)
- **PANEL 1:** Applications of AI/ML in mass movement hazard mapping
  - Deep learning for mapping retrogressive thaw slumps and landslides across the Arctic permafrost domain (*Ingmar Nitze, Alfred Wegener Institute for Polar and Marine Research*)
  - Constructing a large-scale landslide database across heterogeneous environments using task-specific model updates (*Savinay Nagendra, Computer Science, Penn State*)
  - Rapid mapping of event landslides using deep-learning (*Nikhil Prakash, Colorado School of Mines*)
  - Transformation of big imagery into Arctic science-ready products (*remote participation*, *Chandi Witharana, Natural Resources & the Environment, University of Connecticut*)
- **PANEL 2** - Applications of AI/ML in mass movement hazard susceptibility prediction
  - Landslide hazard prediction using ML techniques and physics-enhanced ML (*Te Pei, Civil and Environmental Engineering, Penn State*)
  - Explainable neural network for accurate and interpretable landslide susceptibility modeling: Part I Background and methodology (*Khalid Youssef, University of California - Los Angeles)* Part II Landslide susceptibility application (*Kevin Shao, University of California - Los Angeles*)
- Coffee breaks, lunch, and dinner provided the opportunity for informal networking between attendees.

May 13 - DAY 2: Integrating AI/ML into mass movement hazard in cryosphere

- Summary of Day 1 (*workshop hosts*)
- Understanding permafrost slope dynamics in lowlands with a terrain-cryofacies approach (*remote participation, Eva Stephani, Alaska Science Center U.S. Geological Survey*)
- Influence of subsurface properties in rock slope failure (*remote participation, Louise Vick, Geosciences, University of Tromso*)

- Mapping permafrost thaw-related slope failures in Alaska's Arctic National Parks (*remote participation, David Swanson, National Park Service*)
- Thinking big: Permafrost Discovery Gateway and Arctic T-SLIP (*remote participation, Anna Liljedalh, Woodwell Climate Institute)*
- The 2012 Lituya rock-ice avalanche in Alaska: preliminary insights from remote sensing and dynamic analysis (*Kaushal Gnyawali, University of British Columbia)*
- Closing remarks, next steps, overview of second workshop in 2023 *(Workshop hosts)*

- Coffee breaks and lunch provided the opportunity for informal networking between attendees.

The CryoSlideRisk Workshop 1 convened on Thursday, May 12, 2022. It was held as a hybrid meeting, offering the option for attendance depending on location and COVID-19-related concerns. The following people were in attendance (31 total participants):

Workshop co-hosts: Tong Qiu and Chaopeng Shen (Penn State), Margaret Darrow and Louise Farquahrson (UAF)

In person: Darren Beckstrand (Landslide Technology), Rafael Caduff (Gamma), Kaushal Gynawali (UBC), Kaytan Kelkar (UAF), Jiangtao Liu (Penn State), Savinay Napendra (Penn State), Ingmar Nitzke (Alfred Wegener Institute), Te Pei (Penn State), Nikhil Prakash (Colorado School of Mines), Alexandra Runge (Alfred Wegener Institute), Keven Shao (UCLA), John Thornley (Golder), Khalid Youssef (UCLA)

Online: Lukas Arenson (BGC), Nicole Benshoff (NPS), Matt Billings (ADOT&PF), Denny Capps (NPS), Jeff Currey (ADOT&PF), Nicole Guinn (University of Houston), Stephan Gruber (Carleton University), Line Rouyet (NORCE Norwegian Research Centre AS), Eva Stephani (USGS), David Swanson (NPS), Mahendra Udawalpola (University of Connecticut), Louise Vick (Arctic University of Norway), Chandi Witharana (University of Connecticut)

Summary of discussion on overview presentations (*by Chaopeng Shen and Stephan Gruber*)

A lot of what we are trying to understand lurks at depth (literally). A nice summary from Stephan's presentation indicated that climate change is a major driver, ground ice is a key determinant, and understanding processes is key in anticipating where hazards will occur. An important question that arose early on is, "How do we choose what we want to answer?" This question was raised by others in the discussion, identifying that we need to have data to help us prioritize what areas are fixed first. We are also trying to address concerns and changes that we (modern people) have not experienced in the past. The past is no longer the answer to the future. An additional complexity is that ground temperatures differ from air temperatures; there is a lag that depends on the amount of ice in the ground, so we cannot directly link air temperatures to ground processes. We also do not really know where the ground ice is, or what type of ice is present. Do we already have mathematical equations that describe these hazards, or are there parameters/processes/pieces of the process that we just do not know? We need to use machine learning (ML), which is based on previous patterns, to address permafrost degradation, which is currently occurring along novel trajectories. This will require many discussions to build the understanding necessary between permafrost scientists/engineers and the ML community. This will take time and good communication.

Key takeaways: In order to address infrastructure adaptation due to mass-movement hazards amid climate change in the cryosphere, we need to predict a sequence of processes: atmospheric warming, permafrost thaw, mass-movement risk, and vulnerability of infrastructure, where our ability to predict each process is dependent on the accuracy/reliability of our prediction of the previous process. ML can help with these predictions; however, we need to build adequate understanding between permafrost researchers, practitioners, policy makers, and ML specialists. Dialogs and workshops like this will be helpful and needed.

Summary of Panel I discussion *(presentations by Ingmar Nitze, Savinay Nagendra, Nikhil Prakash, Chandi Witharana - focused on Applications of AI/ML in mass movement hazard mapping)*

Challenges raised in this group of presentations included delineating mass movement extents through ML. Training to determine linear features is good to delineate headwalls, but perhaps it could be done in reverse by delineating the polygon, and then finding the highest point as the head scarp. We also need good quality spatial data at a high temporal resolution, as things are constantly changing. In images, cloud cover presents challenges but can be masked out, this just takes additional steps in the workflow.

Challenges arise due to different geographic regions having different image quality, different vegetation cover, different topography, etc., making it difficult to apply one workflow to multiple regions or areas. To be more robust, ML needs to incorporate training data from a wide range of locations. The lack of vegetation in a retrogressive thaw slump (RTS) is an example of how the characteristics of landslide features can vary, even within a single feature. The head scarp has a lack of vegetation and a sharp topographic expression, but on the lower body, the vegetation grows back more quickly. One area of interest is to develop an algorithm to analyze the vegetative differences on the landslide body to define the edges. One problem with this is that these features are small and this spatial variability may not be identifiable in more widely accessible satellite imagery of a medium to low (e.g., >3m) resolution.

A key issue that arose is that models trained on only one region are not transferable. Members of the workshop group presented and discussed different ML approaches, and cost/time savings associated with each. Some deep learning (DL) models identify clusters of features as one feature, missing the parts that separate them. Mapping features from more than one year is valuable, to see change with time. Practitioners indicated that these applications will be valuable for mapping and selecting future infrastructure corridors.

What we discussed at this workshop is not limited to the cryosphere. ML can be applied to disaster relief efforts (e.g., Prakash's presentation). It is important to understand the scale of the disaster; hence, fast mapping is critical, even with sacrificing certain accuracy. When a disaster occurs, even if people living in the area have the expertise to help with the relief, they are also personally affected by the disaster, and may not have the infrastructure (such as electricity, lodging) to do the necessary analysis. Mapping should be done by an outside group. To help others, there should be a global network that updates a global inventory week by week. Take an area of seismically-triggered landslides - are they all dangerous? We may want to focus on those that block valleys because of breakout flooding, or those that block roads because they impede relief efforts. To determine this, responders need a hotspot map or a risk map, etc. There needs to be large-scale automated mapping to avoid the need to do this processing at critical times. Some data sources are restrictive; not all countries have access to the data. A remote sensing group that assists first responders, like the GEER group, could help with the response work.

Key takeaways: Using ML to map mass-movement hazards has great potential but also comes with challenges associated with the lack of a global (or even regional) landslide databases (which are expensive and time consuming to build), unique landslide spatial features, and poor transferability of ML models trained from an ecoregion or a mass-movement type to another; however, reliable and accurate maps for future mass-movement hazards will be useful for practitioners and policy makers for infrastructure planning purposes.

Summary of Panel II discussion *(presentations by Te Pei, Khalid Youssef, Kevin Shao - Applications of AI/ML in mass movement hazard susceptibility prediction)*

Incorporating geologic/engineering parameters and knowledge into ML models is valuable, yet difficult to obtain and often regionally inconsistent. Models can and should be modified based on expert opinion and domain knowledge. Which type of model or approach is the right one? Are some better for some applications? Do we try several and compare the results? Are there time constraints or constraints based on computing power? Physics-guided machine learning (PGML) is an emerging paradigm in ML. It aims to leverage the complementary strength between domain knowledge and the power of data science. PGML is promising as a linkage between permafrost and ML communities in integrating permafrost knowledge into ML algorithms to improve the prediction of atmospheric warming, permafrost thaw, mass-movement risk, and vulnerability of infrastructure.

Summary of general group discussion at end of Day 1

- Existing geohazard maps are great (albeit, large scale), but they take time to prepare and they are static. How do we modify those to account for a changing climate?
- At the site-specific scale, analysis involves putting instruments in the ground to create a precise design mitigation. Applying ML to large-scale problems is a struggle, but adding in geologic parameters has promise. There are different geo-datasets from geologic mapping, including soil surveys that have near-surface soil distribution including strengths or other properties.
- Risk means different things to different people. Is a landslide that is moving in the middle of nowhere a human problem? From an infrastructure perspective, it can be reduced to dollars. Are we going to have to plan for it? Will there be impacts due to climate and permafrost degradation?
- When we look at mass movements, are they gravity-driven processes or phase-driven processes? What do we understand, and how do we classify them? This is important

when considering ML. We need to extract resilience based on changes in probability and how climate change is affecting probability of failure.

- As an overview of ML, there are purely data-driven models - if you have a large amount of data, these are the best-performing models. It is possible to add some additional constraints, such as the soil/rock properties. In this case, the model uses fundamental aspects of machine learning, resulting in a process-based assessment. Philosophically, do you think the laws are true? If there is unknown knowledge, then this differential approach is the way to fill in gaps, train with less data than the purely data-driven models. We can pose a question, assuming that the other parts are relatively robust, so that we know what to fill in and what to learn. For mass movement, the community of practitioners can provide data sets or parameters that contribute to slope failure. Can ML be used to identify those data gaps? Can we use ML results as a tool to go back into the field to measure something that we are now overlooking?
- Constraining the model and the process is the way to go. A long-term goal may be to incorporate domain knowledge from different data sets and drive forward. With computational processes, this would incorporate domain knowledge from all groups.
- One of the issues is that we do not have enough data, because we typically only look at hazards that really affect the infrastructure corridors. If we work with a larger data set that incorporates areas away from communities and infrastructure, this may help us to determine the different mechanisms of failure.
- Typically, field studies are constrained to local areas. ML may help us grow that data base and expand the knowledge on what is causing slope failures. The field investigators are focused on subsurface properties. How can DL techniques use areas where we have this information, and then expand into other areas using remote sensing data?
- Data limitations are huge: we have only rudimentary ground ice maps; little understanding of ground ice distribution in mountainous environments, type of ice (i.e., pore ice, massive ice), and a lack of understanding of the permafrost degradation mechanisms in mountainous environments. This is an issue because ice is often a key element in slope stability in permafrost-affected regions and when it melts it can drastically weaken slope strength. We sometimes find ice where we do not expect it; perhaps we do not completely understand the processes that formed that ice, or the processes that operated in the past are quite different than current conditions (e.g., an avalanche that occurred long ago and was buried with talus, then becomes unstable in a changing environment). Perhaps formation of ice in the lowlands or buried glacial ice is easier to understand; mountain permafrost is difficult to predict. There are legacy effects. Can ML help to fill in the gaps? There are issues with data availability in the cryosphere. We have only coarse data for first-order estimates. This will be a challenge with modeling. Another challenge is wildfires – they expose more soil to the atmosphere, and then more carbon is released, adding to the feedback loop. Mountain permafrost in the Himalaya gets less attention than other permafrost locations. This will also degrade and cause slope failures and contamination of reservoirs. Can ML predict this? Perhaps at this workshop, this is the first group to discuss this type of convergence.

- This situation is not unique to geohazards; it is similar in hydrology. On the positive side, so many other things that are correlated with the soil properties. We do not have to understand it completely because of data synergy. If we compile a lot of data across the entire terrain, then we get something that is more robust. There are more commonalities than dissimilarities; a variety of data could be helpful.
- Let's say that we have a model; how do you test it? The quantity of data itself is not the only problem. If we have a few measurements that are spread around, this is better for training. For ML, we could go to a site to measure all those processes, but also doing something at a global scale gives us better understanding. When we look at global patterns, we may be able to see answers that we cannot at the regional scale. We need automated techniques to look at the global scale.
- When we talk about mass movement in the cryosphere, this is a sensitive environment. Things are dynamic with time. Most of the models ignore this. There needs to be an emphasis on the timestamp and how long it takes the system to recover from the impact. It is critical to include changes with time. For permafrost, we could start by looking at changes in Alaska from 1950 to 2000, to see if we can identify the lag effect and if specific areas of the landscape are more susceptible.
- We are doing pretty well with landslide mapping. We can greatly increase the extent of the dataset. There is so much to learn from the data. Now we need to call ML and permafrost experts and develop synergy. Right now, the applications are so far apart from each other that it is hard to compare and know what is correct. We need to arrive at a common ground so that everyone understands… the processes that lead to the patterns ML predicts.
- Let's dream big. Practitioners want to have those maps and scroll through time to see where the hazards are increasing. This can help to answer this dilemma: when you have funds, where should you invest your limited funds? Which is the most critical location that requires your investment? Geohazard management is a dream product.
- Those in the ML community say that data are the problem and wonder how to scale better. How are we able to have our community grow together and not individually? One example of success is RTS mapping. We have already started to have these datasets for benchmarking and training, have thought about how to build the data set and how to standardize the digitizing. Then moving forward, all groups will have the same process. Another example is in the hydrology realm. There is a data set (CAMELS) with different attributes. It has become a benchmark for people to test. Folks in other countries started to notice this, and started to have a common data format. One benefit for the work involved in producing the data set is that all authors get acknowledged and have high citations. Researchers are out there collecting data – we can provide them with standardized processes, and they can provide the information. If we start somewhere and use one data set, when it gets popular, others are going to use it. We can start with one format, then bring in people from different regions to check it to ensure it works in different topography/vegetation, etc. (much like the IPA working group for mapping rock glaciers).
- Any DL group focused on this area can start with a low resolution image; with DL, we can make it have super resolution, but we need data to train it and a standardized way to

train it. The general idea is that if the data are interpolated and increased in resolution, features that are missed are then detected if it is up-scaled. This is not "faking" high-resolution data; instead there is ML to apply it to low-resolution data. It is a good technique to avoid using commercial data. This also could be used to standardize features among different maps/data sets.

- For slope stability analysis, it is 100% site-specific for cohesion and angle of internal friction. We take samples from the failure zone, do residual shear strength tests, then put the results into a stability model and back-calculate strength. This generalizes the results a bit more. We do the same thing with pore water pressure. We install instruments into the ground and measure water and slope movement, and correlate to what triggering factors may exist. We find that the rate of movement has to do more with the rate of change in water level rather than the water level itself. This is something that is not typically modeled.
- Perhaps we develop synthetic data sets for landslides, with susceptibility based only on physics; but we currently do not have enough understanding/rendering of 3-D models to do ML. Such models do exist for rivers and how they change with time. That may be the way to go in the future. Having these very detailed simulations can generate possible scenarios, and then establish trigger levels.
- NEXT STEPS: We need to develop a framework with what the correct milestones should be or stages necessary to develop the "dream big." We can take what is needed from the field-based community as inputs, and then develop a map on where to go next. What is our strategy for the next steps? Depends on science questions and/ or land management issues at hand!

The CryoSlideRisk Workshop 1 continued on Friday, May 13, 2022. The following people were in attendance on Day 2 (28 total participants):

Workshop co-hosts: Tong Qiu and Chaopeng Shen (Penn State), Margaret Darrow and Louise Farquahrson (UAF)

In person: Darren Beckstrand (Landslide Technology), Rafael Caduff (Gamma), Kaushal Gynawali (UBC), Kaytan Kelkar (UAF), Ingmar Nitzke (Alfred Wegener Institute), Te Pei (Penn State), Nikhil Prakash (Colorado School of Mines), Alexandra Runge (Alfred Wegener Institute), Keven Shao (UCLA), John Thornley (Golder), Khalid Youssef (UCLA)

Online: Lukas Arenson (BGC), Nicole Benshoff (NPS), Denny Capps (NPS), Jeff Currey (ADOT&PF), Andy Garrigus (Golder), Nicole Guinn (University of Houston), Stephan Gruber (Carleton University), Anna Liljedahl (Woodwell Climate Institute), Xiaofeng Liu (Penn State), Line Rouyet ((NORCE Norwegian Research Centre AS), Eva Stephani (USGS), David Swanson (NPS), Louise Vick (Arctic University of Norway),

After a quick summary of Day 1, with a check on accuracy by the workshop participants, the workshop continued with five presentations. The presentations were focused more on characterizing permafrost in different regions, such as lowlands and mountainous regions. The following is a summary of that conversation and points raised.

The presentations of ice-rich permafrost in lowland areas and slides in bedrock in mountainous areas demonstrated different layers with different material properties or ice content. The ML community said that this aspect was highly relevant in understanding why some slopes failed and not others. The group pondered if strength properties were tested as a function of temperature. The presentations illustrated the difficulty in obtaining such data because of inaccessibility and cost associated with sampling. Failure in mountainous regions was also attributed not just to the direct thaw of permafrost/melting of ice, but also to melting of snow and movement of the snow melt through the subsurface, as well as the surface being frozen, causing high pore pressures. Suction also forms as the surface is frozen. Thawing in the spring causes pore pressures to build up again. In the winter, even if the rock mass deforms, there is suction holding the rock mass back, artificially increasing the factor of safety.

Desk-top analysis can be done before going to the field to identify areas of interest. All layers describing the surface, including snow distribution, water distribution, vegetation including shrubs, should be evaluated. If the data sets existed, and methods also existed to combine all of those features to identify problem areas, this would be useful. Perhaps we could use ML to map our cores better to identify the cryostructure.

The trigger of widespread events was discussed, whether it is gradual warming, or event-driven, such as heavy rain or snow melt. It is a little of both, although there are little site-specific data for these remote areas to answer this question. Some shapefiles do exist that can be used as training sets for ML.

No matter how we train a model, there will be false positives and negatives. We need experts to interact with those labels to fix them. We need a system where experts can interact with these data and improve the results. IPA action groups exist for developing protocols to develop community-wide training data sets for mapping RTS and rock glaciers. Something similar should be developed for mass movements in general, to develop the same terminology, to exchange ideas, and to bring in people who are non-experts but would like to learn. We can have an interface that allows a user interested in a certain area to choose “show me all thaw slumps,” click on a box, and then the latest model that is pre-trained shows all the thaw slumps. The interface/web page can have an option where one can choose “I want to work on this region” and gradually work to correct labels (false positives/negatives).

One case study that described an example of modeling was exciting from a ML perspective. It represented a start to get a large-scale example that could be used to train a model to create synthetic data. This could generate new and large amounts of data. We could generate new scenarios that do not yet exist. This is an exciting prospect for forecasting how things will change in the future. For events with long runouts, we need some data to calibrate other events. If we have a source, then we can determine how far it is going to move. With ML, we can calibrate the runout.

The workshop ended with the workshop hosts providing an evaluation form for feedback, and an overview of upcoming tasks, such as writing this summary report and collaborating on a "perspectives" paper.