

# AI/ML Meeting Notes

Meeting summaries and links in reverse chronological order.

## June 8, 2021 - Using Deep learning to Identify North American Bats

**Recording:** On [Microsoft Stream](#) for Dept of Interior employees.

**Abstract:** The North American Bat Monitoring Program (NABat) is a multi-agency coordinated monitoring program that uses standardized protocols to gather data on the 47 bat species in the U.S. and Canada. Bats are notoriously difficult to monitor, however some monitoring methods, such as passive acoustic sensors, have seen success in recent years. As of 2021, NABat has collected over 8.7 million acoustic samples of bats from a wide range of habitats. Using this data, we have begun developing an open source reference library and deep learning environment to better serve the conservation needs of the bat community.

**Bio:** Ben Gotthold is a Computer Scientist with the North American Bat Monitoring Program in Fort Collins, Colorado. Ben has worked to design and build the cloud based technology stack that serves the diverse needs of this large scale monitoring effort.

## 2021 April 13 - Turning nature into numbers: Advancing Ecoscience through the lens of remote sensing and artificial intelligence/machine learning

**Title:** Turning nature into numbers: Advancing Ecoscience through the lens of remote sensing and artificial intelligence/machine learning

**Recording:** On [Microsoft Stream](#) for Dept of Interior employees

**Abstract:** The EcoAI project supports USGS science initiatives within dryland and high-latitude ecosystems through the integration of field observations, remote sensing, and artificial intelligence/machine learning (AI/ML). Dr. Neal Pastick will discuss how EcoAI makes use of cloud and high-performance computing resources (e.g., USGS HPCs, Amazon Web Services, Google Cloud Platform) and AI/ML to solve complex environmental problems. Specifically, this talk will focus on the USGS Rangeland Exotic Plant Monitoring System that was developed to fingerprint the abundance, spread, and impacts of invasive plants in the western United States. This system currently leverages Harmonized Landsat and Sentinel-2 data and machine and deep learning techniques to assess the past, present, and future abundance of exotic annual grasses and the resistance of sagebrush ecosystems to invasion. If time allows, Dr. Pastick will also present research aimed at better understanding high-latitude ecosystems and the development of a global lake monitoring system for assessing changes in surface area, color, and temperature using Alaska as a pilot study area.

**Bios:** **Neal Pastick** is a Research Physical Scientist with the U.S. Geological Survey (USGS) – Earth Resources Observation and Science (EROS) Center and a Ph.D. graduate from the Natural Resources Science and Management Program at the University of Minnesota – Twin Cities. He has conducted remote sensing research for environmental science applications over the past 11 years, with an emphasis on characterizing changing aquatic and terrestrial landscapes through the assimilation of field studies, remote sensing, and AI/ML.

## 2021 March 9 - Integrating Deep Learning Methods into a National High-Resolution Land Cover Mapping

**Recording:** On [Microsoft Stream](#) for Dept of Interior employees

**Title:** **Integrating Deep Learning Methods into a National High-Resolution Land Cover Mapping**

**Nate Herold** is a Physical Scientist with NOAA's Office for Coastal Management (OCM) where he manages the Coastal Change Analysis Program (C-CAP), NOAA's national land cover mapping and monitoring effort. He has been with OCM for 18 years, is intimately involved with Data and Tool offerings within our Digital Coast site, and is currently the chair of the Multi-Resolution Land Characteristics (MRLC) Consortium, helping to coordinate the land cover mapping work from all major federal agencies.

**Chris Robinson** is a Senior Remote Sensing Analyst with LynkerTechnologies on contract at the NOAA Office for Coastal Management (OCM) where he serves as the lead technical analyst for the Coastal Change Analysis Program (C-CAP). Chris has over 20 years of experience working with high-resolution imagery for the development of coastal land cover and habitat data.

## 2021 February 9 - Assessing the Artificial Intelligence (AI) Readiness of USGS Data

**Recording:** Dept of Interior employees with access to O365 can view the video on Stream [here](#). Or download mp4: [210209-aiml-hou.mp4](#).

**Title:** **Assessing the Artificial Intelligence (AI) Readiness of USGS Data**

This presentation provides an overview of a pilot program at the USGS for understanding the “state of the data” or data maturity as a result of institutional, system, and data level policies and practices. The pilot program includes a component for evaluating specific data characteristics for supporting AI applications. The presentation will share the initial AI readiness evaluation approach and the preliminary results.

Sophie Hou, Contractor to USGS, CSS Science Data Management Branch

## 2020 November 10 - Semantics & machine reasoning: the (other) AI road to EarthMAP?

**Recording:** Dept of Interior employees with access to O365 can view the video on Stream [here](#). Or download mp4: [201110-aiml-bagstad.mp4](#)

**Title: Semantics & machine reasoning: the (other) AI road to EarthMAP?**

**Abstract:** Despite widespread growth in open data and machine learning, substantial challenges remain to the reusability and interoperability of scientific data and models. Since 2007, the Artificial Intelligence for Ecosystem Services (ARIES) project has been developing infrastructure for integrated, multidisciplinary scientific modeling using two AI tools – semantics and machine reasoning. These automate the assembly of multidisciplinary scientific data and models appropriate to the user's context (i.e., location and spatiotemporal scale) of interest. Semantics apply consistent terminology to data and model components, enabling a computer system to recognize compatible data/model elements. Interdisciplinary semantics are particularly challenging to develop and apply, but ARIES has demonstrated that robust, modular, interdisciplinary semantics are possible. Machine reasoning enables a computer system to make choices when presented with alternative options – i.e., to use a particular model or dataset in a given application. A semantic web system like ARIES provides an environment for scientists to add new data and models to a global ecosystem for coupling, testing, adjusting, and reusing models – in particular, specifying appropriate conditions for model reuse. At the same time, a simple web interface provides access to data and models for a location and time period of interest, enabling non-technical users (like DOI resource managers) to run models, explore results and management tradeoffs, and view full model provenance. ARIES has been used to address diverse scientific and natural resource management questions globally. Although substantial work remains to achieve large-scale application, ARIES' underlying technology may provide inspiration to what an integrated, AI-enabled system like EarthMAP could achieve.

**Bio:** Ken Bagstad is a Research Economist in the Geosciences & Environmental Change Science Center in Denver. His research interests span the modeling and valuation of ecosystem services, bridging the worlds of economic and natural capital accounting, and ecoinformatics. Since 2007 he has been actively involved in the Artificial Intelligence for Ecosystem Services (ARIES) project, an international collaboration to build a semantic web application supporting networked, automated multidisciplinary modeling for decision making.

## 2020 October 13 - Injecting process knowledge into neural networks for more accurate predictions

**Recording:** [201013-cdi-aiml.mp4](#)

**Title: Injecting process knowledge into neural networks for more accurate predictions**

**Abstract:** We have applied Process-Guided Deep Learning (PGDL) to water temperature prediction in several recent studies, supporting fisheries assessment in hundreds of lakes in the Upper Midwest and informing timed releases of cold water from reservoirs into streams of the Delaware River Basin. Our PGDL models, which integrate physics knowledge into neural networks, outperform baseline deep-learning and process-based models with respect to prediction accuracy and reliable detection of threshold exceedances. A rapidly growing community is applying similar methods to modeling tasks in other fields, from climate to translational biology, and the approach holds promise for numerous USGS-relevant applications. In this talk I will dive into the details of the neural network structures, physical constraints, and training methods that are responsible for the success of PGDL models to date.

**Bio:** Alison Appling has been a water data scientist with the US Geological Survey since 2015. She has a bachelor's degree in Symbolic Systems from Stanford University and a PhD in Ecology from Duke University. Her research addresses the movement of energy, carbon, and nutrients through rivers, lakes, and floodplains, with an emphasis on using data science and machine learning to improve the estimation and prediction of water quality variables.

## 2020 September 8 - Advancing AI-ML for the USGS in the CHS environment

**Recording:** [200908-AIML-recording.mp4](#) (People with access to the Microsoft Team can also stream the recording from the Recordings tab, or go to the GS-CDI Channel)

**Slides:** [200908-AIML-kuckuk.pdf](#)

**Abstract:** The Cloud Hosting Solutions (CHS) program is now offering and actively supporting the utilization of various artificial intelligence and machine learning (AI-ML) services. Matt Kuckuk will describe the kinds of support that are or will be provided to CHS customers. Matt will describe his recommendations for how investigators can identify use cases that are most likely to benefit from application of AI-ML techniques, and how they can begin to determine what standard algorithms to evaluate. He'll discuss, for example, how "scientific" use cases and "operational" use cases differ in terms of their requirements. Finally, he will describe how to engage with the CHS AI-ML team to get support for new proposed use cases and applications.

Matt Kuckuk recently joined the CHS team after decades leading AI, ML and data analytics practice teams of up to 200 data scientists and developers in large consulting companies. He has implemented a wide variety of AI-ML applications and research projects for public sector as well as commercial organizations. He is now focused on creating and sustaining the AI-ML capability within CHS to advance the USGS mission.

## 2020 July 14 - New York Water Science Center AI and Gage-Cam

Gage-Cam is a low cost, custom built wireless web camera paired with a custom deep learning algorithm that allows for a computer vision method to measure water surface elevation (stage). This project is a joint venture between Web Informatics and Mapping and the New York Water Science Center. Today's topic will be a short presentation on Gage-cam's design, capabilities, and prototyping. This will be followed by an open forum discussion on the technology and engineering behind the sensor, emerging methods in AI and single board "Lite-Tech" based devices researched by NYWSC-AI.

Daniel Beckman holds a bachelors degree from the University of Colorado in Ecology and Evolutionary Biology and minors in Chemistry and Computer Science. He currently attends Graduate school at the University of Colorado School of Engineering and Applied Science where he studies Machine Learning and Artificial Intelligence. He has worked in data for almost two decades in a variety of fields including, counterintelligence, research & development, forensic chemistry, and genomics. Daniel joined the USGS in 2017 and WIM in 2018. Currently, he works in cloud integration.

Recording: [200714-AIML-recording.mp4](#) (People with access to the Microsoft Team can also stream the recording from the Recordings tab, or go to the GS-CDI Channel)

Slides: [200714-beckman-AIML.pdf](#)

## 2020 June 9 - USGS Tallgrass Supercomputer 101 for AI/ML

Recording: [200609-AIML-recording.mp4](#)

Slides: [200609-AIML\\_Meeting\\_Rapstine.pdf](#)

Natalya Rapstine will give an overview of new USGS Tallgrass supercomputer designed to support machine learning and deep learning workflows at scale and deep learning software and tools for data science workflows.

Natalya Rapstine is a computational scientist in the [Science Analytics and Synthesis \(SAS\)](#), [Advanced Research Computing \(ARC\)](#) group of the Core Science Systems. She has a bachelor degree in Earth Science from Rice University and a MS in Statistics from Colorado School of Mines, and she has been with the USGS since 2016. Her expertise is in high performance computing, machine and deep learning applications for advancement of science at U. S. Geological Survey.

## 2020 May 12 - Amazon SageMaker

Recording: [2020-05-12-CDI-AIML-recording.mp4](#)

Speaker bios:

Inseok Heo is a data scientist in Envision Engineering AWS.

Inseok received a PhD from the department of Electrical Engineering in the University of Wisconsin Madison in 2015. He specializes in speech and audio signal processing and machine learning. In his career, he developed and worked on single/multi channel noise reduction, beamforming, and Alexa wakeword recognition/detection for Amazon Echo device.

Amogh Gaikwad is a Solutions Architect, specializing in AI/ML, for AWS Federal customers and is part of the specialist team for Analytics. Prior to his role at AWS, Amogh has worked as a software developer, developing enterprise applications. Through his role at AWS his has created ML solutions to help federal customers migrate their AI/ML workloads to AWS.

Amogh has received his Master's Degree in Computer Science specializing in Big Data Analytics and Machine Learning from George Mason University

Phillip Dawson is a geophysicist with the U.S. Geological Survey's Volcano Science Center, focusing on theoretical and experimental investigations of active volcanism and volcanic processes. He currently works on the Seismology of Magmatic Injection project at the California Volcano Observatory, Menlo Park, California. This project is dedicated to understanding the underlying physics driving volcanic seismicity and processes through the use of detailed field experiments and the application, modification, and extension of existing seismic methods and theories.

## 2020 April 14 - Fine scale mapping of water features at the national scale using machine learning analysis of high-resolution satellite images: Application of the new AI-ML natural resource software - DELTA

- [2020-04-14-CDI-AIML-recording.mp4](#)

Michael Furlong, NASA-Ames Intelligent Robotics Group  
Jack Eggleston, USGS WMA Hydrologic Remote Sensing Branch  
John Stock, USGS Innovation Center

**Abstract:** The availability of high-resolution satellite imagery, combined with machine learning analysis to rapidly process the satellite imagery, provides the USGS with a new capability to map natural resources at the national scale. The new capability is made possible by technology progress in these areas:

- 1 - Daily national imagery at <1 to 5 m pixel size from commercial providers
- 2 - High-performance computing (USGS high-performance computing or Cloud)
- 3 - Artificial intelligence and machine learning (AI-ML) tools to automatically process the imagery

USGS is working to build enterprise capability in each of these 3 areas and has a growing focus on development of AI-ML tools. In this presentation, two USGS projects that rely on collaborations with external partners to develop AI/ML tools to map water extent will be discussed. In one of these projects USGS is collaborating with the NASA-Ames Intelligent Robotics Group to use its Deep Earth Learning Training, and Analysis (DELTA) software. The DELTA software will be presented including description of its early implementation on the USGS TallGrass supercomputing system.

## 2019 November 12 - AI/ML in USGS enabled by Tallgrass: classifying golden eagle behavior using telemetry

Log in to access this month's recording and slides.

**Abstract:** This presentation provides an overview of how we use a recurrent autoencoder neural network to encode sequential California golden eagle telemetry data. The encoding is followed by an unsupervised clustering technique, Deep Embedded Clustering (DEC), to iteratively cluster the data into a chosen number of behavior classes. We apply the method to simulated movement data sets and telemetry data for a Golden Eagle. The DEC achieves better unsupervised clustering accuracy scores for the simulated datasets as compared to the baseline K-means clustering result.

**Speaker Bio:** Natalya Rapstine is a Computer Scientist at Advanced Research Computing group, specializing in computational data science, statistics, and machine learning applications for advancement of science at the U.S. Geological Survey. She received a M.S. in Statistics from Colorado School of Mines.

## 2019 October 8 - Radiant MLHub: A Repository for Machine Learning Ready Geospatial Training Data

- [2019-10-08\\_CDI-AIML-recording.mp4](#)

**Abstract:** Satellite observations provide invaluable data across different spatio-temporal scales. These data enable us to build models for applications such as land cover classification, agricultural monitoring, surface water mapping, biodiversity monitoring, among others. Meanwhile, machine learning (ML) techniques can be utilized to advance these applications, and develop faster, more efficient and scalable models. These techniques learn from training datasets that are generated from image annotation or ground reference observations. However, to develop accurate ML-based models, and be able to validate their accuracy, we need to use benchmark training datasets that are representative of the diversity of the target variable, and openly accessible to all researchers and developers.

To address this requirement, Radiant Earth Foundation has established Radiant MLHub to foster sharing of geospatial training data for different thematic applications. Radiant MLHub is hosted on the cloud and users will be able to search for different training datasets, and quickly ingest them into their pipelines using an API. To increase interoperability of training datasets generated by different institutions, Radiant MLHub has adopted the SpatioTemporal Asset Catalog (STAC) as the standard for data cataloging.

In this presentation, I will review the architecture of Radiant MLHub, its API access and the STAC definition for training data. Next, I will present two applications on using ML models for LC classification from multi-spectral data and surface water detection from Synthetic Aperture (SAR) data.

**Speaker bio:** Hamed Alemohammad is the Chief Data Scientist at Radiant Earth Foundation, leading the development of Radiant MLHub as an open source cloud native commons for Machine Learning applications using Earth Observations. He has extensive expertise in machine learning, remote sensing and imagery techniques particularly in developing new algorithms for multi-spectral satellite and airborne based observations. He also serves as an elected member of the American Geophysical Union's technical committee on remote sensing. Prior to joining Radiant Earth, he was a Research Scientist at Columbia University. Hamed received his PhD in Civil and Environmental Engineering from MIT in 2014.

## 2019 September 10 - Utilizing Deep Neural Networks for Landscape Conservation: An Application of Google's Tensorflow for a Cannabis Production Inventory in Northern California

- Speaker: Daryl Van Dyke, spatial analyst for the USFWS.
- [Recording for 2019-09-10](#)

**Abstract:** Landscape classification is the task of using imagery to map defined features on the landscape. As computer technology and data science methodology advances, new techniques for this problem emerge. Modern machine learning (ML) utilizing neural networks (NN) – is becoming an industry-standard data science approach for a variety of applications. In particular, procedures of analysis for the task of computer vision (CV) are particularly adept and well-understood at the task of computer vision (CV).

However, current landscape classification necessarily exposes trade-offs between accuracy, spatial granularity, and resources required. CV offers a unique combination of speed and accuracy, while still producing feature mappings rather than simple pixel classifications. Compared to other explicit feature extractors, such as object-based image analysis (OBIA), CV can be a cost-effective a powerful methodology for obtaining features from difficult-to-classify image domains.

This application shows how an off-the-shelf Deep Neural Network (DNN) algorithm – Inception v2 – was retrained into a production classifier and applied to the problem of locating and sizing cannabis production on private lands in Trinity County. This application demonstrates the strengths and limitations for applying this method at the landscape scale.

The presentation concludes with 'next steps' and identifies developing technologies and architectures that mitigate some of the limitations in the current application.

**Speaker bio:** Daryl Van Dyke serves as the spatial analyst for the USFWS, in Science Applications and Strategic Habitat Conservation. I have a interdisciplinary background, with a focus on community and environment as well as second BS and MS in Environmental Engineering. My thesis focused on using two-dimensional hydrodynamics for fish passage culvert retrofit design. As a federal servant, and a programmer, I've looked at the developing technologies of LiDAR, Structure-from-Motion, and ML as pivotal to the task of landscape analysis and conservation design. Non-technical interests in federal service include integrating analytic workflows, encouraging cross-program collaboration, and building accountability and reproducible science in resource management.

## 2019 August 13 - Continuous streamflow and nearshore wave monitoring from time-lapse cameras using deep neural networks.

- Daniel Buscombe (Northern Arizona University) described a proof-of-concept study into designing and implementing a single deep learning framework that can be used for both stream gaging and wave gauging from appropriate time-series of imagery.
- [Recording for 2019-08-13 Part 1](#)
- [Recording for 2019-08-13 Part 2](#)

**Abstract:** The expense and logistics of monitoring streamflow (e.g. stage and discharge) and nearshore waves (e.g. height and period) using in situ instrumentation such as current meters, bubblers, pressure transducers, etc, limits the extent to which such important basic information can be acquired. Machine learning might offer a solution, if such information can be obtained remotely from time-lapse imagery using inexpensive consumable camera installations. To that end, I describe a proof-of-concept study into designing and implementing a single deep learning framework that can be used for both stream gaging and wave gauging from appropriate time-series of imagery. I show that it is possible to train the framework to estimate 1) stage and/or discharge from oblique imagery of streams at USGS gaging stations, using existing time-lapse camera infrastructure; and 2) nearshore wave height and period from oblique and rectified imagery from USGS Argus systems. This proof-of-concept technique is based on deep convolutional neural networks (CNNs), which are deep learning models for regression tasks based on automated image feature extraction. The stream/wave gauge model framework consists of an existing generic CNN model to extract features from imagery - called a 'base model', with additional layers to distill the feature information into lower dimensional spaces, prevent overfitting, and a final layer of dense neurons to predict continuously varying quantities. Given enough training data, the model can generalize well to a site despite variation in, for example, lighting, weather, snow cover, vegetation, and any transient objects in the scene. This development might offer the potential to train models for imagery at sites based on short deployments of in situ instrumentation, especially useful for sites where instrumentation is difficult or expensive to maintain for long periods. This entirely data-driven technique, at least for now, must be trained separately for each site and quantity, so would be suitable for very long-term, site-specific estimation of wave or hydraulic parameters from stationary camera installations, subsequent to a training period. Further development might promote low-cost (or even hobbyist) hydrodynamic and hydraulic monitoring anywhere.

## 2019 June 11 - Strategic Science Planning at USGS

- Pete Doucette provided a review of recent Strategic Science Planning at USGS. This included thoughts captured from the 21st Century Science Update Workshop at NCTC, and the CDI Workshop in Boulder, CO, held May and June 2019.
- [Recording for 2019-06-11](#)

## 2019 May 14 - XGBoost in Continuous Change Detection and Classification (CCDC); Deep learning to quantify benthic habitat

- Announcements (Pete Doucette)
  - Pete and other members of an AI/ML focus group presented to the USGS Executive Leadership Team at the end of March. Associate Directors are enthusiastic about incorporating AI/ML into their mission area research.
  - Don't forget the CDI workshop is happening June 4-7, 2019 in Boulder Colorado, tomorrow, May 15 is the last day for registration. Virtual participation links will be posted on the [Workshop Page](#) by the week before the workshop.
  - The AI/ML group leads are still interested in collecting an inventory of your AI/ML projects to share with interested USGS leadership. [You can fill out the form here.](#)
- XGBoost in Continuous Change Detection and Classification (CCDC) - Chris Barber, USGS EROS
- Deep learning to quantify benthic habitat - Peter Esselman, USGS Great Lakes Science Center
- [Recording for 2019-05-14](#)

## 2019 April 9 - no meeting

## 2019 March 12 - Infrastructure for Deep Learning at the USGS; AI for Ecosystem Services

- Jeff Falgout presented on Tallgrass, a new machine for AI at the USGS
- Ken Bagstad presented on AI for integrated environmental modeling & forecasting (+ overview of AI for Ecosystem Services)
- JC pointed out some activity on the [AI/ML forum](#) and encouraged members to post
- Group leads reminded members to contribute to a [spreadsheet for collecting USGS AI/ML project descriptions](#) to communicate to USGS leadership
- [Recording for 2019-03-12](#)

## 2019 February 12 - Innovation Center Opportunities and AI and Land Imaging

- John Stock talked about opportunities at the USGS Innovation Center related to AI/ML, including [postdoctoral positions](#)
- Pete Doucette gave a presentation “Ruminations on AI and Land Imaging,” covering some background to artificial intelligence and machine learning, relevant Landsat and Analysis Ready Data activities at the USGS, and the importance of team science
- See [February CDI collaboration area blog post](#) that summarizes the call
- Group leads asked members to contribute to a [spreadsheet for collecting USGS AI/ML project descriptions](#) to communicate to USGS leadership
- [Recording for 2019-02-12](#)

## 2018 December 11 - Inaugural Meeting

- Introduction to the group by Tim Quinn, Chief, Office of Enterprise Information
- Introduction to the wiki space and AI/ML forum, JC Nelson
- Comments from attendees, including mention of [dl\\_tools toolbox for deep learning](#), an output from a recent CDI funded project
- [Full notes from JC Nelson](#)
- [Recording for 2018-12-11](#)