

Introduction:

- Satellite image time-series for monitoring Earth system change (land surface reflectance)
- Harmonic models for sequentially extracting seasonal parameters with Particle Filtering
- state vector: $x_{b,k} = [\mu_{b,k}, \alpha_{b,k}, \varphi_{b,k}, \omega_{b,k}]$ (mean, amplitude, phase, frequency at k in band b)
- **Multispectral behavior at change point**

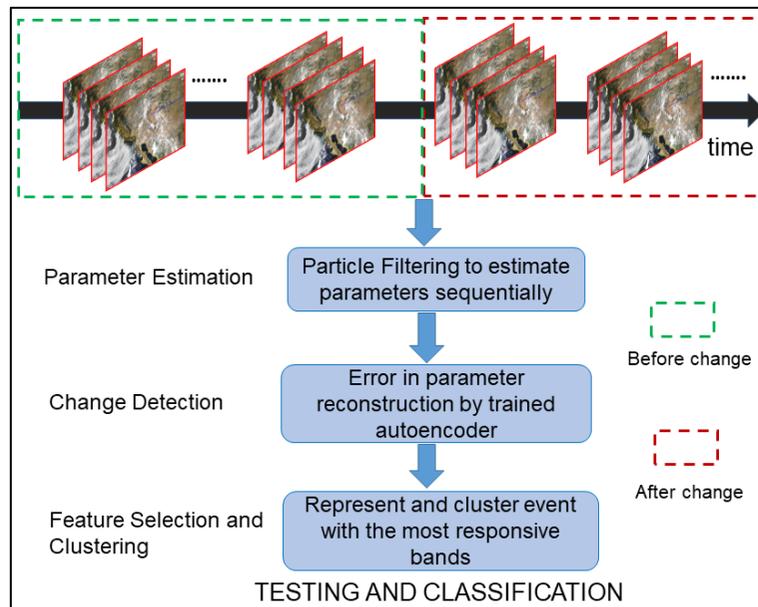
Methodology:

Dataset: MODIS land surface reflectance (7 bands, 8 days, 500 m)

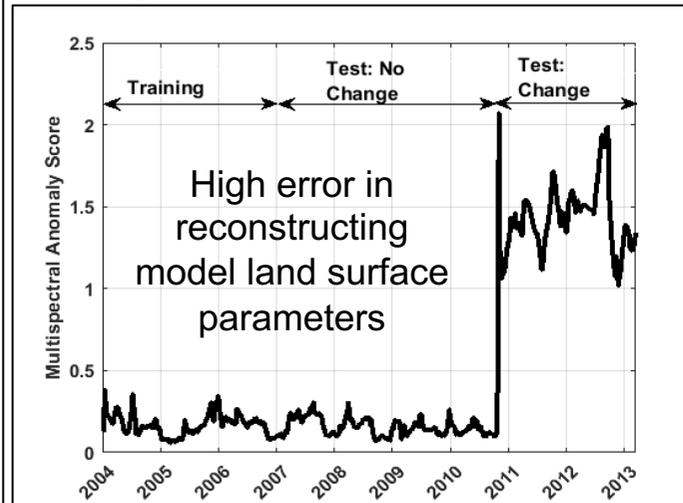
Pixel time-series of regions over 16 years

Change Events: forest fire, drought, flood (coastal wetland, agricultural area), coastal land gain

Training: pre-change spectra of land surface reflectance time series across all bands to learn expected spectral reflectance model using autoencoders



Change Detection:



Sequentially Estimated reflectance: $(x_{b,k})$
Reconstructed reflectance: $(x'_{b,k})$

Anomaly Score:

$$A_k = \sum_{b=1}^B |x'_{b,k} - x_{b,k}|$$

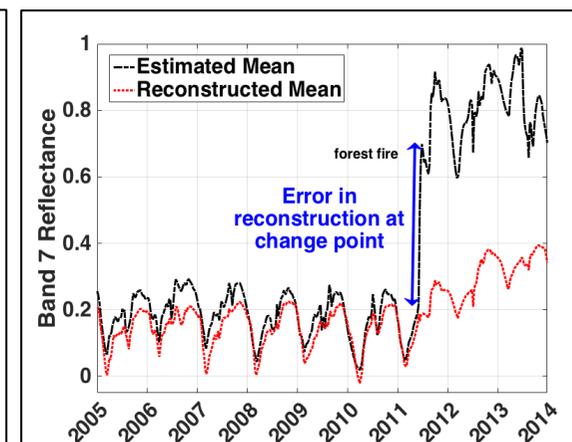
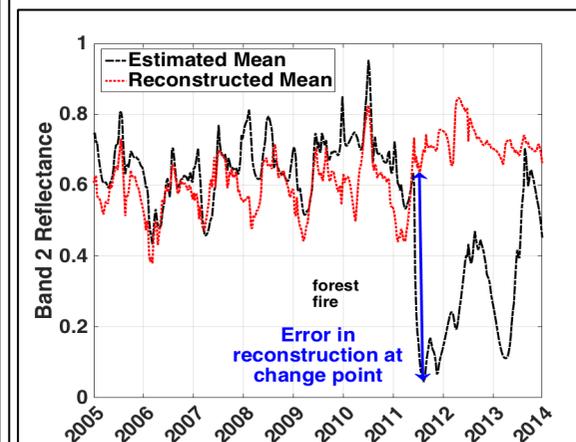
Interpretability of Reconstruction Errors:

$$(x'_{b,k} - x_{b,k}) > 0$$

Time-series (at k in band b) **decreases** due to change

$$(x'_{b,k} - x_{b,k}) < 0$$

Time-series (at k in band b) **increases** due to change

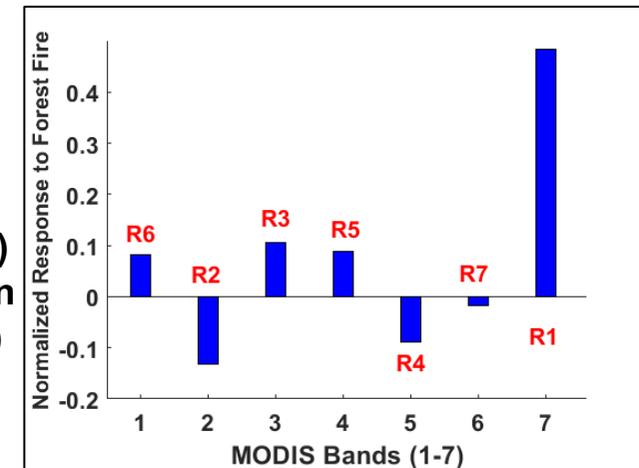


Reconstruction error over time for forest fire in NIR and SWIR

Change Signatures:

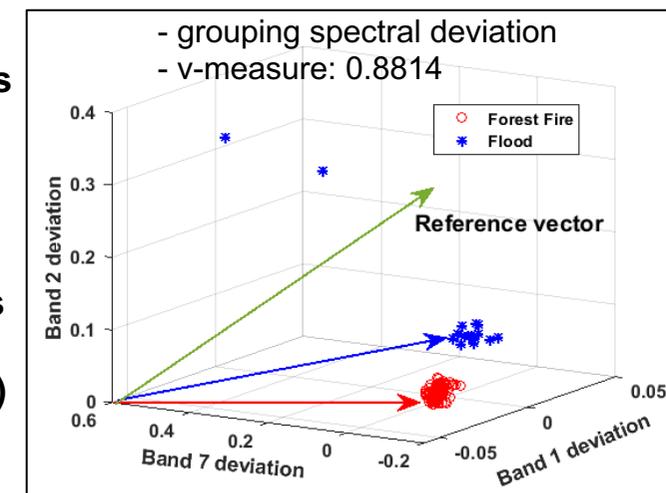
Similarity with NDVI:

- Increase in band 1 (Red)
- Decrease in band 2 (NIR)



Clustering:

Uniqueness of change vector of different events (deviations of change signatures)



Conclusion and Future Work:

- Unique signature of each event from deviations
- extension to more regions, hyperspectral data

References:

1. S. Chakraborty, et al. "Time-varying modeling of land cover change dynamics due to forest fires." *IEEE JSTARS*, 11.6 (2018).
2. S.Chakraborty, A.Papandreou-Suppappola, P.R.Christensen, "Class Separability of Land Cover Change Events from Multispectral Satellite Image Time-Series." *AGUFM* 2019: IN43A-05.

Prediction of Global Geomagnetic Field Disturbances using Recurrent Neural Network

Hyunju Connor (hkconnor@Alaska.edu)¹, Shishar Priyadarshi¹, Matthew Blandin¹, and Amy Keesee²
¹University of Alaska Fairbanks, ²University of New Hampshire

MAGICIAN Team for Forecasting GICs

MAGICIAN Team



MAGICIAN is a joint UAF – UNH team funded by the 2018 NSF EPSCoR RII Track 2 Program that develops machine-learning algorithms for predicting hazardous Geomagnetically Induced Currents (GIC).

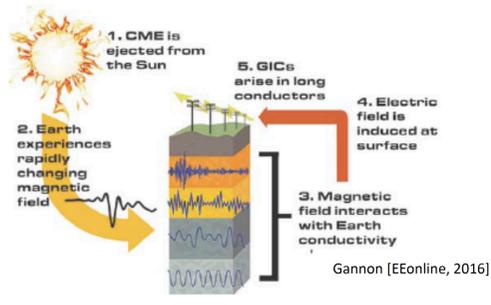
- Check other presentations of MAGICIAN team at this AI & Data Science Workshop!
 - Poster #53: Using an LSTM and Classification Methods to Determine Risk of dB/dt Threshold Crossings as Proxy for Geomagnetically Induced Currents - Michael Coughlan, UNH.
 - Poster #57: Comparison of Time Series Techniques to Model Connections Between Solar Wind Input and Geomagnetically Induced Currents - Amy Keesee, UNH.

Meet MAGICIAN Team!



Motivation

Forecast of Geomagnetically Induced Currents (GICs)



Gannon [EOnline, 2016]

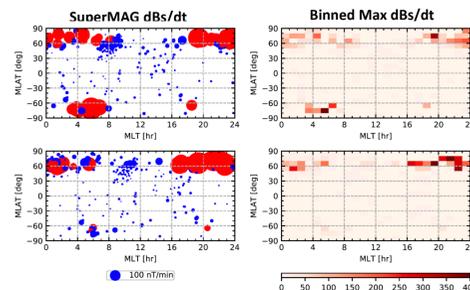
- Worldwide/Nationwide GIC data are not available.
- Our study focuses on the prediction of large geomagnetic field disturbances, a trigger of GICs.

Multi-variate LSTM Model for dB/dt Prediction

Recurrent Neural Network (RNN) for predicting geomagnetic field disturbance (dB/dt)

- DATA:**
 - OMNI solar wind and IMF conditions in 2012 and 2015
 - SuperMAG surface & vertical disturbances (i.e., dBs/dt & dBz/dt) in 2012 and 2015
 - 80% for training, 20% for validation, and the 2012-03-09 storm for testing
- Method:** Multi-variate Long Short Term Memory (LSTM) network
 - 50 neurons in a single hidden layer, 50 epochs with a batch size of 72
 - Adam's stochastic gradient descent as an optimization algorithm
 - Mean absolute error as a loss function
- Two machines are trained for dBs/dt and dBz/dt predictions.**
 - Input: IMF Bz, Solar Wind Density, dBs/dt (or dBz/dt) at a previous minute (t-1)
 - Output: dBs/dt (or dBz/dt) at the next minute (t)

SuperMAG Data Binning

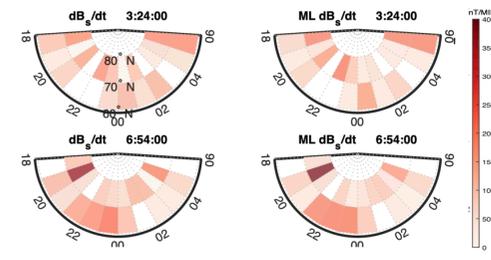


- To provide an even spatial resolution, we binned SuperMAG data into 10° MLAT x 1hr MLT grids and select max dBs/dt and max dBz/dt in each bin as our dataset.
- Red circles indicate potential GIC locations where dB/dt went higher than 100nT/min in the past few hours.

Improvement needed in future

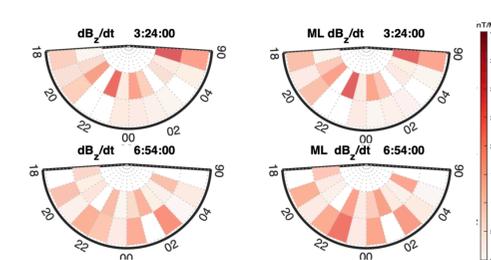
- Train machines with longer periods of data.
- Use a longer time history of input (e.g. 60 mins of SW/IMF data).
 - Remove dBs/dt and dBz/dt from the input.
 - Our current model may find stronger correlation with dB/dt at t-1min.
- Consider sophisticated solar wind propagation from the bow shock to each bins.
 - 60min delay from the bow shock to the nightside bins were assumed.
- Consider finer spatial resolution for higher latitude
- Use better validation techniques than RMSE [Welling et al. SW2018; Maimaiti et al. SW2019; Camporeale. JGR 2020]
- Use different machine learning techniques
 - Multi-layered LSTM, Artificial Neural Network, Convolutional Neural Network, Principal Component Analysis, etc.

Model-Data comparison of dBs/dt over the Nightside Northern Hemisphere



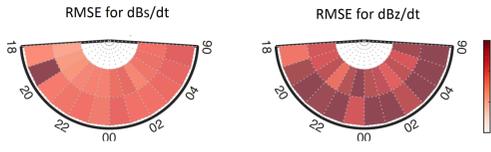
Our machine-learned predictions (right) show a good agreement with the binned SuperMAG data (left) on a larger spatial scale.

Model-Data comparison of dBz/dt over the Nightside Northern Hemisphere



Our machine-learned predictions (right) show a good agreement with the binned SuperMAG data (left) on a larger spatial scale.

Model Results over the Nightside Northern Hemisphere



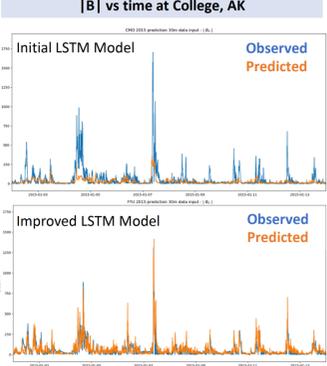
Considering $\pm 30\text{nT/min}$ of RMSE, our machine may miss or falsely predict $\sim 100\text{nT/min}$ of geomagnetic disturbance.

However, it won't be troublesome to forecast several hundreds nT/min that potentially produces a catastrophic GIC event.

Other MAGICIAN Team Activities

MAGICIAN Team Activity 1: Multi-variate LSTM model for AK array

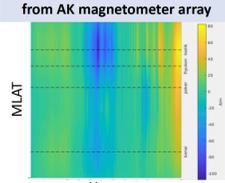
|B| vs time at College, AK



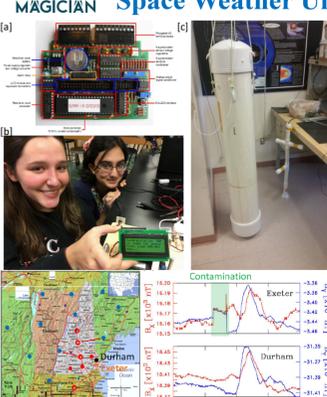
We are developing/improving a LSTM model for the AK magnetometer array by using various input parameters and tuning LSTM parameters/functions.

Once matured, the model can predict the ionospheric currents above AK as well as the GIC risk in AK.

Ionosphere currents predicted from AK magnetometer array



MAGICIAN Team Activity 2: Space Weather UnderGround Project



- UNH SWUG (Chuck Smith, lead) is expanding to UAF (Don Hampton, lead).
- Undergrad and high school students assemble and deploys a low-cost, but research-capable fluxgate magnetometer (1sec and 1nT resolution) across AK and New England.
- MAGICIAN team will gather the magnetometer data in higher spatial resolution and used them for the GIC studies.

MAGICIAN Team Activity 3: Pathfinder GIC Project with GVEA



- Golden Valley Electric Association (GVEA) is the biggest power company in Alaska.
- MOU between UAF and GVEA is on progress to measure GICs for the next 5 years.
- As our first attempt, GIC will be measured at two locations. Multiple candidate locations are currently discussed.
- GIC data will be available to the public.

Summary

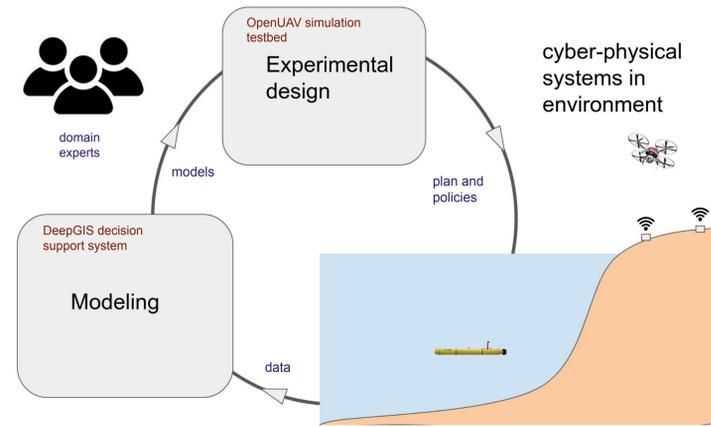
- We developed a prototype of a multi-variate LSTM model using 2 years of OMNI and SuperMAG data.
- The prototype model catches over 100nT/min of dB/dt relatively well on 09 Mar 2012 geomagnetic storm.
- Once matured, this model can provide an advanced warning of GICs that are typically triggered by large dB/dt.
- In addition to the ML-GIC models, MAGICIAN team provides the low-cost, research-capable magnetometer arrays in AK and NH, and the GIC measurements in AK to the space science community.

This work is supported by NSF EPSCoR Grant #1920965.

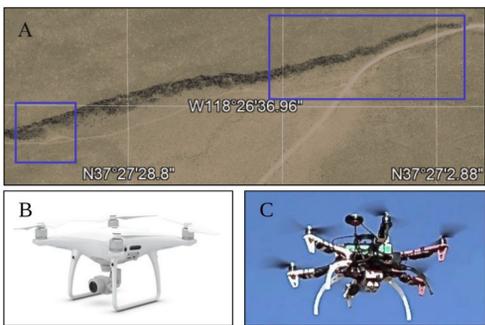
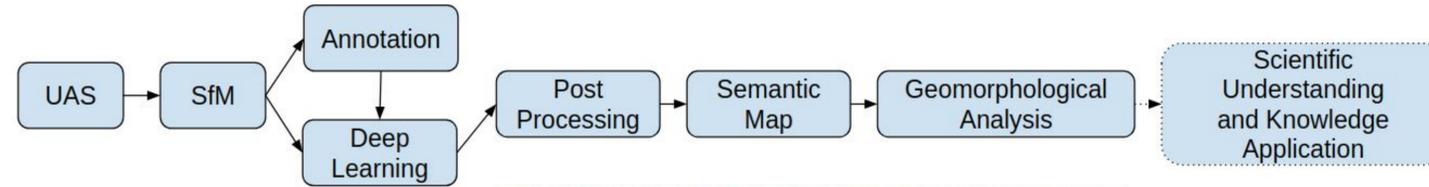
The Annotation Game: Towards Collaborative Science with Humans, Robots, and AI

Poster #3

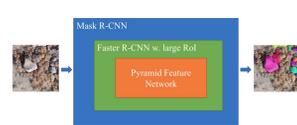
Zhiang Chen, Tyler R. Scott, Ethan Duncan, Harish Anand, A.L.G. Prasad, Sarah Bearman, Devin Keating, Chelsea Scott, Brent Hayashi, Mark Wronkiewicz, Jnaneshwar Das, Ramon Arrowsmith



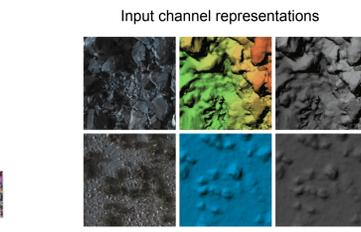
Data-driven rock geomorphology with UAS



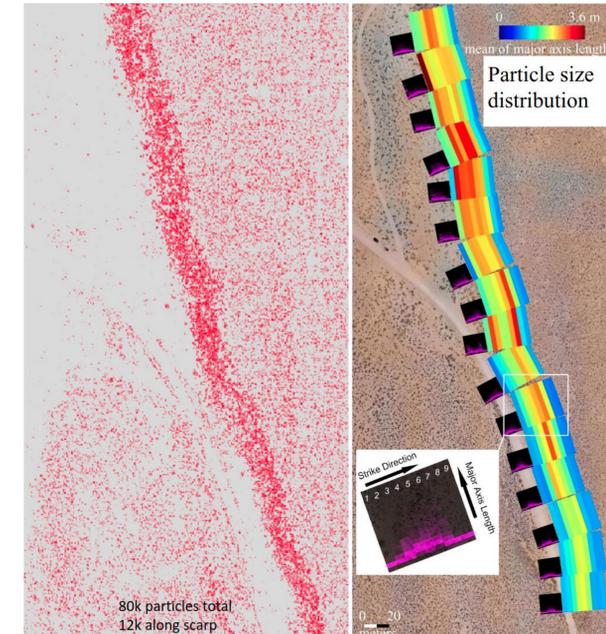
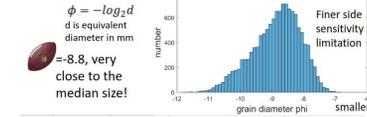
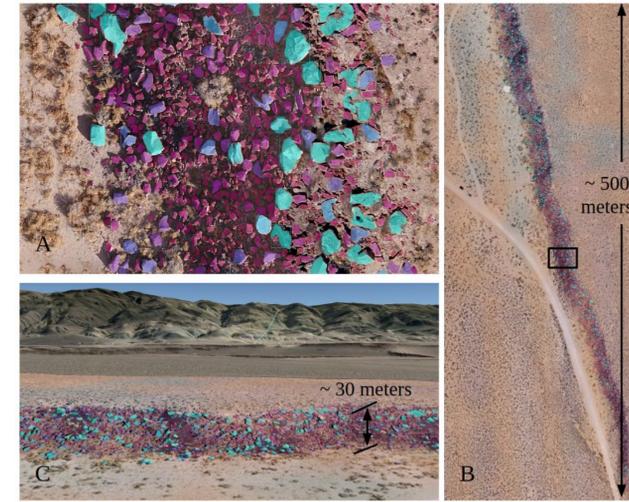
Deep neural network architecture



- Transfer learning > no transfer learning
- RGB + DEM3/DEM1 > RGB
- DEM1 alone is not enough due to bushes
- RGB + RE + NIR > RGB
- RGB + RG + NIR + DEM3/DEM1 have the best performance



- Mask R-CNN: Instance segmentation (individual rocks) up to 200 rocks in an input image
- Faster R-CNN with large number of Regions of Interest: rock size varies largely in an input image (major-axis length: 0.2 ~ 3.6 meters)

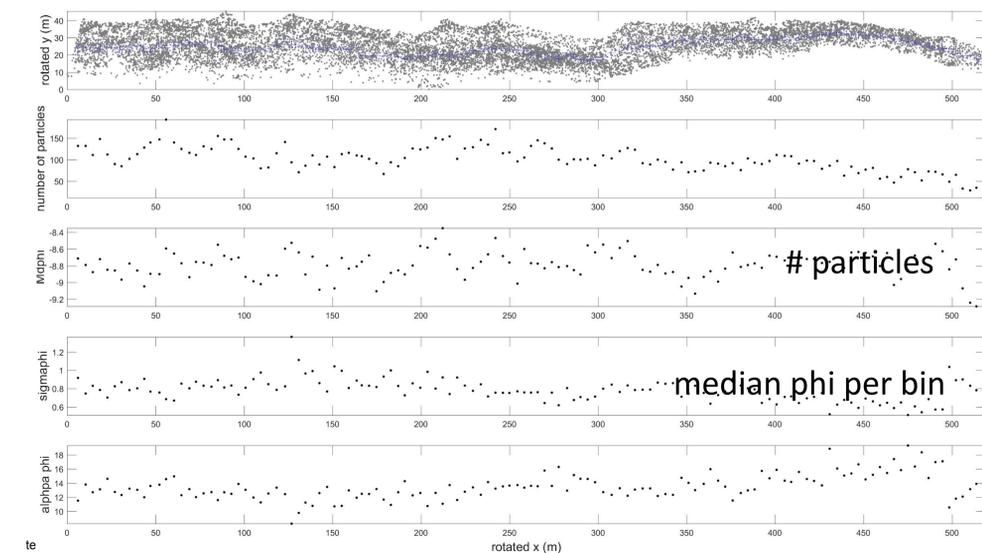
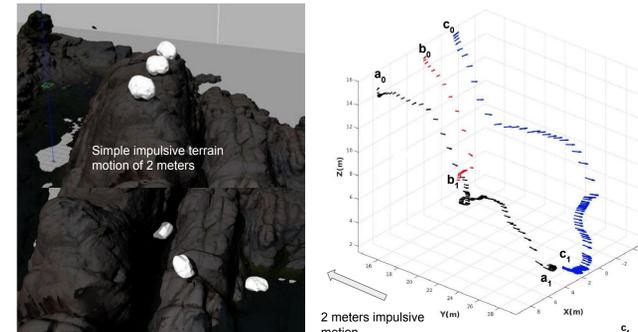
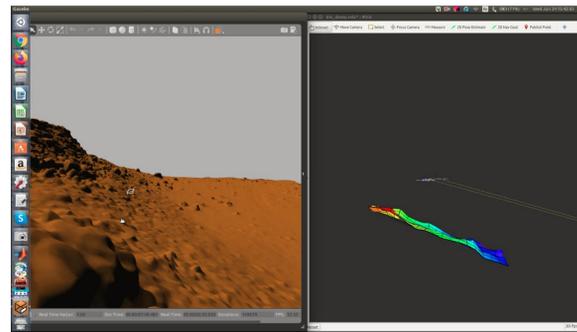


Terrain relative navigation for UAS through trajectory optimization

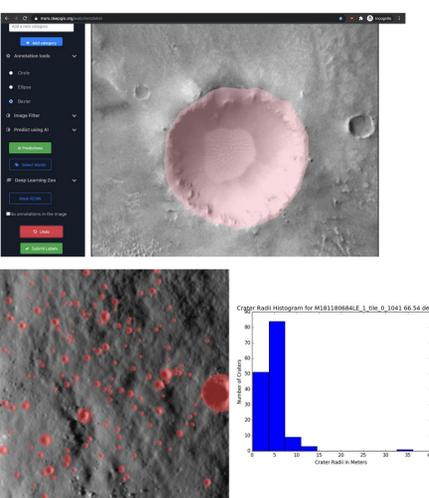
Rock particles can be placed and manipulated in mapped terrain

Fragile geologic feature, trajectory analysis

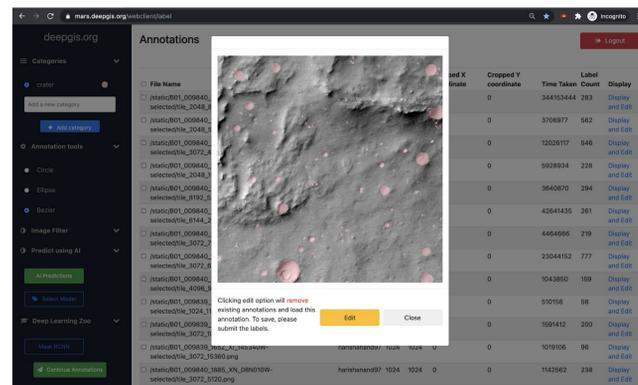
Mission planning and Monte Carlo simulations in OpenUAV



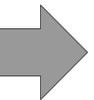
Planetary scale crater mapping



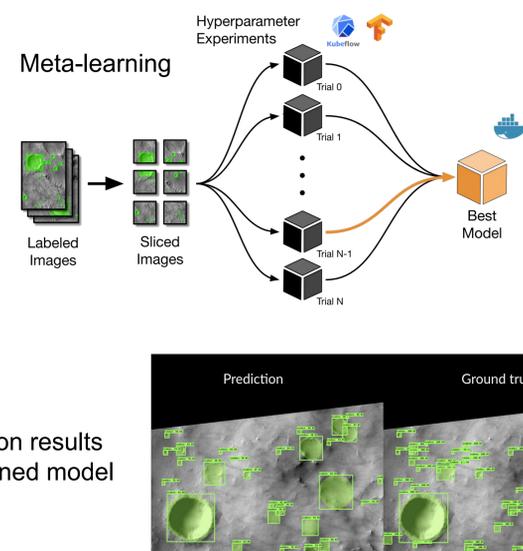
Annotation and analytics with DeepGIS



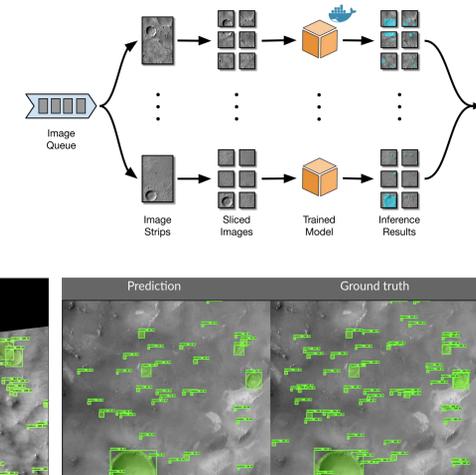
Quality controlled annotations



Prediction results with trained model



Large scale inference on Google Cloud Platform



Summary

- Build upon robotic sampling, cloud, and machine learning.
- Leverage Bayesian optimization for metalearning, to guide annotation, model retraining, and large scale inference.
- AI assisted annotation and model retraining, lifelong learning with experts in the loop.

Acknowledgements: This work was supported in part by Southern California Earthquake Center (SCEC) award 19179, National Science Foundation award CNS-1521617, and National Aeronautics and Space Administration STTR award 19-1-T4.01-2855.

References:
1. Zhiang Chen, Tyler R. Scott, Sarah Bearman, Harish Anand, Chelsea Scott, J Ramon Arrowsmith, Jnaneshwar Das, "Geomorphological Analysis Using Unpiloted Aircraft Systems, Structure from Motion, and Deep Learning", in at IEEERSJ International Conference on Intelligent Robots and Systems (2020)
2. Harish Anand, Jnaneshwar Das, Zhiang Chen, "The OpenUAV Swarm Simulation Testbed: a Collaborative Design Studio for Field Robotics", submitted to IEEE International Conference on Robotics and Automation, 2021.



development SEED

ASU Arizona State University



SC/EC Southern California Earthquake Center





Automatic Per-Pixel Classification of UAVSAR Imagery for Hurricane Flood Detection

Second AI and Data Science Workshop for Earth and Space Sciences, February 9-11, 2021
Poster ID: 4

Authors: Michael Denbina (michael.w.denbina@jpl.caltech.edu),
Zaid J. Towfic, Matthew Thill, Brian Bue, and Yunling Lou
Jet Propulsion Laboratory, California Institute of Technology

1. Introduction and Objectives

- The main objective of this project was to assess the viability of convolutional neural network-based image classifier architectures to automatically detect flooded areas in polarimetric radar imagery collected by the NASA/JPL Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR) instrument. UAVSAR is a fully polarimetric L-band synthetic aperture radar, flown aboard a NASA Gulfstream III aircraft (see Figure 1).
- The objectives of our project were as follows:
 - Build a sizable data set of matched UAVSAR imagery and flood masks, to use as training and testing data for the machine learning algorithms. We are using UAVSAR data collected after Hurricane Harvey for this purpose.
 - Train two pixelwise predictors, U-Net and SegNet (both established convolutional neural network architectures) using the flood labels from objective #1, and evaluate the pixelwise prediction accuracy of each candidate model on blind validation data.



Fig. 1. UAVSAR mounted on a NASA Gulfstream III aircraft. The UAVSAR radar pod is mounted under the body of the aircraft.

2. Automated Flood Mapping Techniques

- There are pre-existing methods for detecting flooding in UAVSAR imagery based on manual interpretation (time-consuming), or simple decision tree or thresholding (low accuracy). However, we believed these methods were likely to be outperformed by more modern image classification approaches, including convolutional neural networks.
- Convolutional neural network architectures such as fully-convolutional-networks and encoder-decoder networks, as shown in Figure 2, immediately lend themselves to our per-pixel labeling effort. These networks utilize only convolutional and other spatial feature-preserving layers throughout the architecture in order to capture discriminative spatial patterns represented in the input images associated with the pixelwise labels.
- We tested two well-known architectures from the literature, SegNet and U-Net, and compared the results to a pre-existing decision tree classifier used as a baseline.

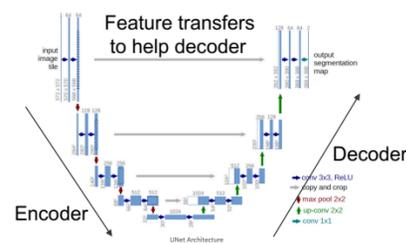


Fig. 2. Illustration of the U-Net architecture, from Ronneberger et al., "U-Net: Convolutional Networks for Biomedical Image Segmentation," MICCAI 2015, <https://arxiv.org/abs/1505.04597>

3. Manual Labelling of Flooded UAVSAR Data

- We developed a simple labelling tool which shows the user an image segment from the UAVSAR imagery, alongside various reference data to aid in determining the appropriate label. The user can click the button corresponding to the desired label for that segment, and the tool will then prompt the user with the next segment. An example is shown in Figure 3.
- We labelled 10873 total image segments (covering over 3.5 million UAVSAR image pixels). After manual labelling, we extracted 40000 randomly located, overlapping 64x64 pixel image patches from two flight lines to use for training. 300 randomly located, non-overlapping patches were extracted for testing/validation and kept separate from the training set.

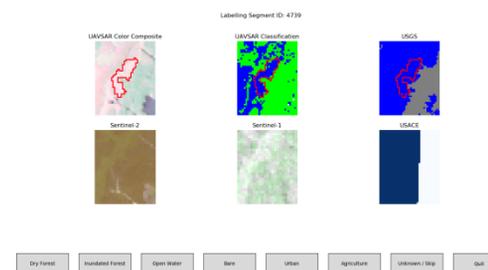


Fig. 3. Example of the UAVSAR manual flood labelling tool, for a flooded forest area.

4. Results and Discussion

- We generated predicted maps of flooded and non-flooded areas for UAVSAR data collected over the Texas Gulf Coast during flooding caused by Hurricane Harvey in August 31 - September 1, 2017, as shown in Figure 4.
- We assessed the classifier accuracy using manually labelled testing data held out from training, as well as NOAA aerial imagery (used for validation manually at specific points shown in Figure 4 as black circles, limited by the cloud-free coverage of the imagery). While the map shows five different classes, for the purposes of assessing the classifier accuracy, we broke the results down into only two classes, flooded (containing open water and flooded vegetation) and non-flooded (containing all other classes).
- The overall accuracy of our U-Net classifier on the manually labelled testing data was 87%, with 85% accuracy for our SegNet classifier. The Kappa coefficient of U-Net was 0.73, with F1 score for the non-flooded class of 0.89 and F1 score for the flooded class of 0.84. SegNet had similar but slightly lower accuracies than U-Net. The baseline classifier (decision tree method) had overall accuracy of 75%, with Kappa coefficient of 0.48, non-flooded F1 score of 0.81, and flooded F1 score of 0.66. **Both U-Net and SegNet outperformed the baseline classifier.**
- Using the NOAA aerial imagery, U-Net had overall accuracy of 82%, with Kappa coefficient of 0.65. SegNet had similar but slightly lower accuracy (by ~2%). None of the classifiers are able to accurately identify flooding in urban areas, however, which is an area of future work (to collect more training data in these spatially complex, heterogeneous environments).

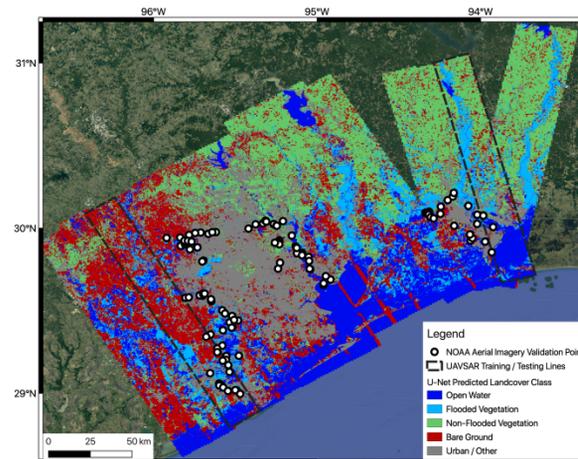


Fig. 4. Map of the study area affected by Hurricane Harvey, showing the U-Net predicted classes generated from the UAVSAR data as input. The dashed black boxes show the UAVSAR flight lines used to collect training data. The black circles show points used to validate the map using NOAA aerial imagery.

5. Conclusions and Future Work

- We demonstrated potential of CNN-based image classifiers for classification of flooded areas in UAVSAR data. With U-Net, 87% overall accuracy using manually labelled testing data, and 82% overall accuracy (outside of urban areas) when validated using NOAA aerial imagery (0.65 Kappa coefficient).
- Future Work:** Apply trained classifiers to Hurricane Florence data (see Figure 5) to test transferability of trained classifiers to other study areas and assess classifier accuracy. **Collect more training and testing data, particularly in urban environments**, which is a current limitation of this approach. Test other input to classifiers, and increased patch size.

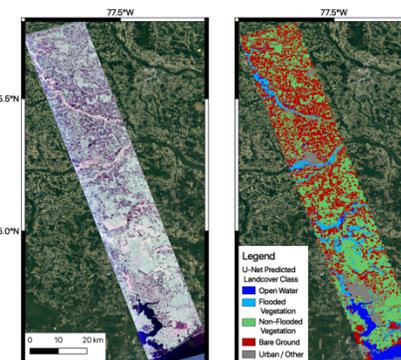
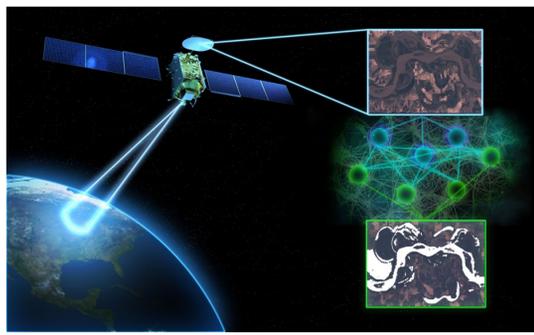


Fig. 5. UAVSAR data (left) and map of predicted classes (right) when applying the Hurricane Harvey trained U-Net classifier to UAVSAR data from Hurricane Florence. The end goal is to have a trained classifier ready to go that can be applied to data from future hurricanes.



Poster ID #6 Deep Earth Learning, Training, and Analysis (DELTA) Automating Machine Learning for Earth Science



P. Michael Furlong^{1,2}, Brian Coltin^{1,2}, Scott McMichael^{1,2}, Tevin Achong, Roberto Campbell, Deanna Flynn, Keanu Nichols, Elizabeth Carter³, Kevin Dobbs⁴, Jonathan Eggleston³, Rachel Sleeter³

NASA Ames Intelligent Robotics Group¹ / KBR Inc.² / USGS³ / NGA⁴ **Contact:** brian.coltin@nasa.gov

Introduction

Machine learning has achieved “human-level” intelligence in tasks ranging from object recognition and speech recognition to mastering the game of Go. However, Earth scientists have yet to fully take advantage of deep learning’s potential. The biggest obstacles are lack of expertise, the high barrier to entry for existing deep learning toolkits, and the intensive computational and data requirements. We are addressing these challenges with DELTA (Deep Earth Learning, Training, and Analysis), a toolkit for Earth scientists and commercial analysts to easily apply deep learning to their own problems.

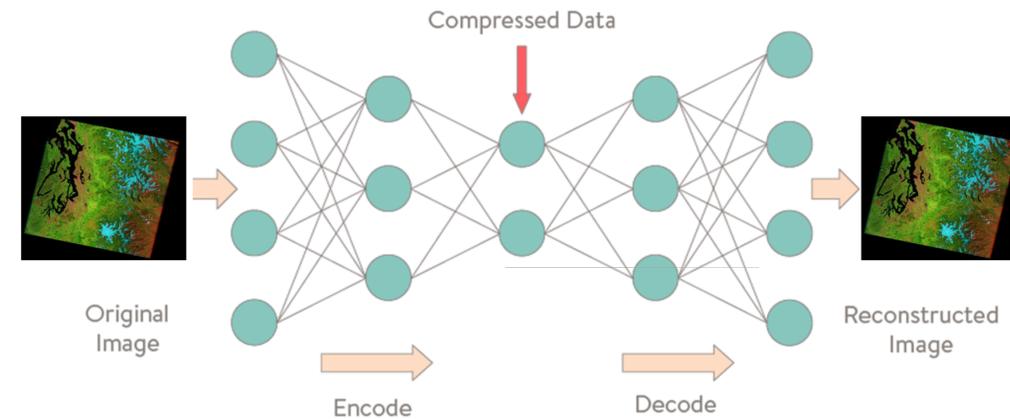
Users only need to provide labeled training data for their problems. DELTA pre-learns useful features for the satellite data sources we are targeting – WorldView, Landsat, and Sentinel-1 -- through large volumes of data on the NASA Pleiades supercomputer. The learned features serve as a starting point to solve users’ specific problems, which can be shared among researchers, amortizing data and computation costs. DELTA then builds task-specific classification networks on top of the learned representations.

We are applying DELTA to mapping flood inundation extent. DELTA will accelerate Earth science research by placing the power of deep learning in the hands of any researcher, achieving “human-level” intelligence in diverse classification tasks which are currently solved manually or with less capable automatic classifiers, drastically expanding the spatial and temporal scales on which many remote sensing problems can be studied.

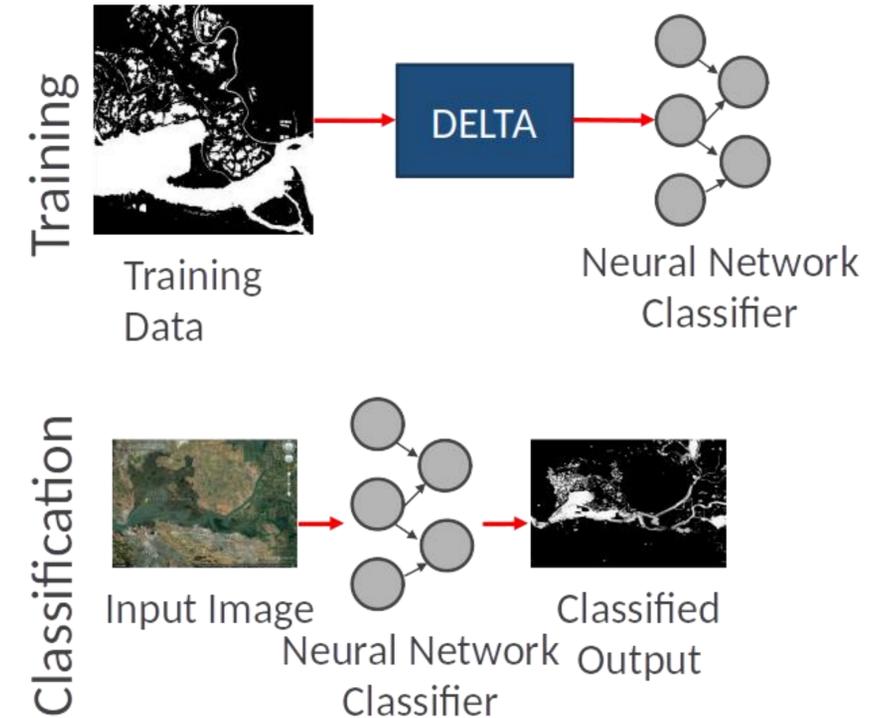
DELTA

- Uses autoencoders to learn features to transform raw satellite data into an informative representation.
- Learns task-specific networks from users' training data, building on the pre-learned features.
- Provides tools for data labeling and visualization of data, natively handles Geotif images.

DELTA is open source and available at <https://github.com/nasa/delta>



Autoencoder: Learn compressed, informative image representation. Different feature sets are learned for our different target satellites, WorldView, LandSat, and Sentinel-1.

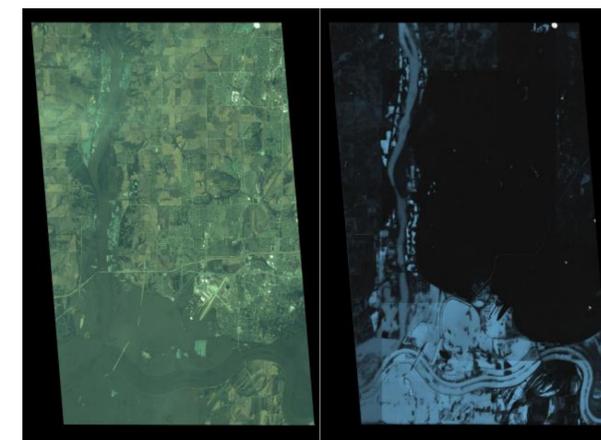


DELTA takes scientists' data and produces a trained neural network. To classify images delta uses the learned network to label new data as part of a larger data processing pipeline.

Results

We have tested our system on Worldview images collected from different flooding events over the past 15 years. Training data was manually annotated by our USGS team members. We trained on 80 images and tested on 20 hold-out images.

Class	Recall	Precision
Water	> 99%	~83%
No Water	~80%	~98%



Left: Input World View flood image (false colour). **Right:** Flood map predicted by DELTA

Acknowledgements

DELTA is a collaboration with the USGS and is funded by the NGA. We also thank our prior collaborators at Google and the USGS in the projects which led to DELTA, especially Josh Livni, Pete Giencke, Marie Pepler, and Brenda Jones.

For more information on our previous flood mapping work, see: *Automatic Boosted Flood Mapping from Satellite Data*. Brian Coltin, Scott McMichael, Trey Smith, and Terrence Fong. International Journal of Remote Sensing, 2016.

Improve Hurricane Intensity Forecast by Machine Learning of NASA Satellite Data

Hui Su¹, Longtao Wu¹, Raksha Pai², Alex Liu³, Peyman Tavallali¹, Albert J. Zhai⁴, Jonathan H. Jiang¹, Mark DeMaria⁵

Poster
#11

¹JPL/Caltech, ²IBM, ³RMDS Lab, ⁴Caltech, ⁵NHC/NOAA
Contact: Hui.Su@jpl.nasa.gov

Objective:

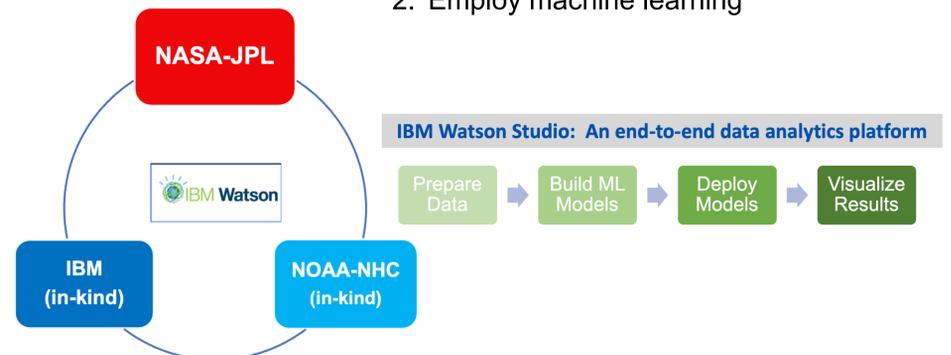
Employ machine learning (ML) techniques and apply NASA satellite observations to improve tropical cyclone (TC) intensity forecast, especially rapid intensification (RI) forecast

Motivation:

- TC intensity forecast has been a challenge for decades
- RI, defined as hurricane maximum sustained wind speed change greater than 30 knots within 24 hours, is particularly difficult to predict. Improving RI forecast accuracy is the top priority of the National Hurricane Center (NHC).
- The NHC's probability of detection (POD) for RI in the Atlantic basin is < 40% and the false alarm ratio (FAR) is > 60% (Kaplan et al. 2015).

Approach:

1. Augment predictors for RI
2. Employ machine learning



Identifying New Predictors for RI

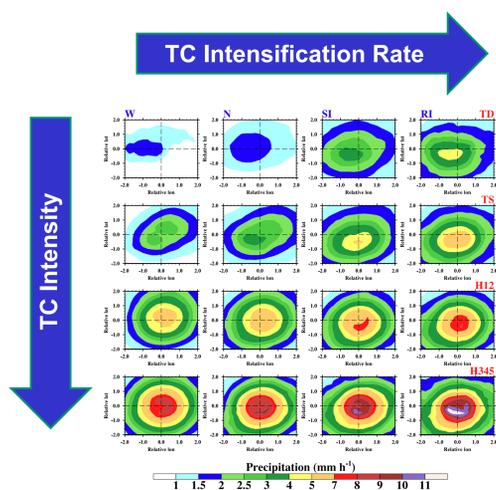


Fig. 1. Composite maps of TRMM precipitation rate in storm-centered coordinate for four TC intensity and four TC Intensification rate groups.

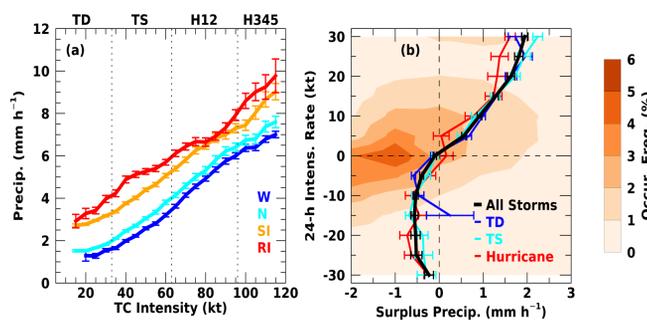


Fig. 2. (a) Composite inner-core precipitation rate as a function of TC intensity for four intensification groups. (b) TC future 24-hr intensity change (DV24) as a function of surplus precipitation for all TCs and three TC intensity groups.

$$\text{Surplus Precipitation } S = P - P_N$$

❖ Inner-core TC precipitation rate, ice water content and outflow temperature bear simple relations with TC intensity change and thus can serve as predictors for RI.

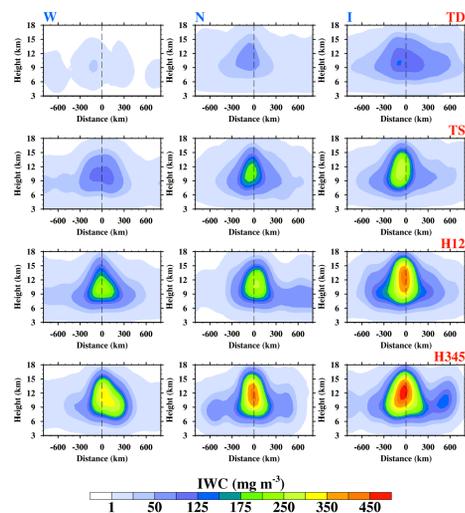


Fig. 3. Composite ice water content profiles in storm-centered coordinate for four TC intensity and three TC intensification rate groups.

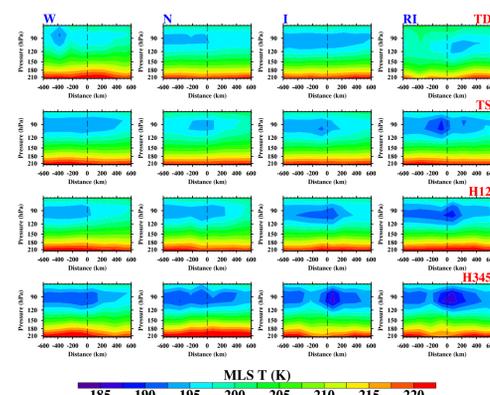


Fig. 4. Composite outflow temperature in storm-centered coordinate for four TC intensity and four TC intensification rate groups.

Building a Machine Learning Model

North Atlantic Basin

○ Training: 2680 cases (1998-2008), Test: 1228 cases (2009-2014)

Eastern North Pacific Basin

○ Training: 2428 cases (1998-2008), Test: 1349 cases (2009-2014)

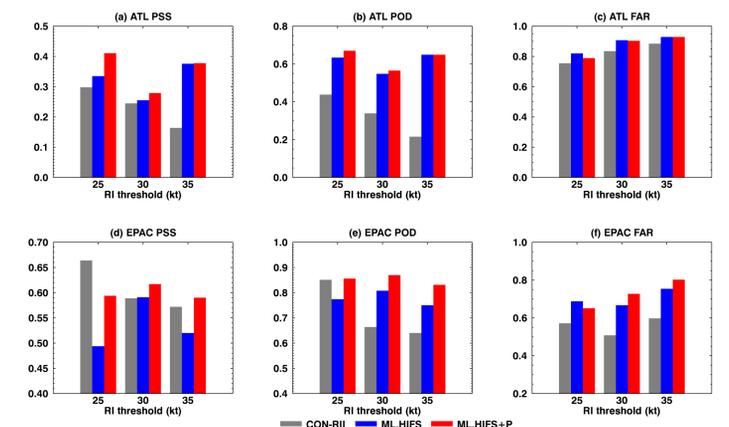


Fig. 5. Predictive skill for RI in North Atlantic (top) and Eastern North Pacific (bottom) for three RI thresholds (DV24 ≥ 25, 30 and 35 kt). (a) and (d) the Peirce Skill Score (PSS), (b) and (e) POD, (c) and (f) FAR. The grey bars are the operational RI consensus forecast scores and the blue (red) bars are the machine learning model using the SHIPS RII predictors without (with) the surplus precipitation from TRMM.

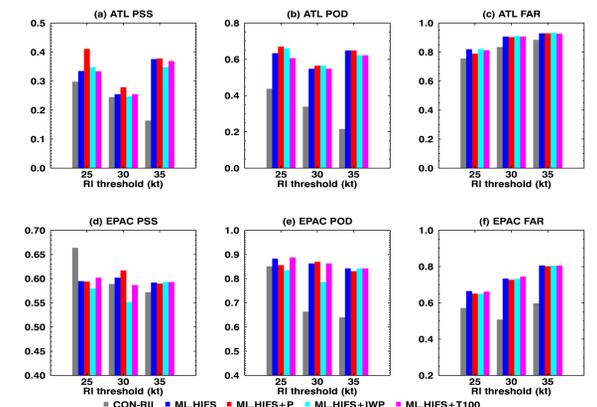


Fig. 6. The forecast scores for RI using the SHIPS RII predictors only, and with the addition of TRMM precipitation (P), MERRA-2 ice water path (IWP) and 100 hPa temperature (T100) as predictors. The grey bars represent the NHC operational RI consensus results.

Conclusions:

- Tropical cyclone intensity change is approximately linearly correlated with surplus inner-core precipitation, ice water path and outflow temperature.
- The JPL-ML model significantly outperforms the NHC operational RI consensus forecast results. Our probability of detection for RI in the Atlantic is 40%, 60% and 200% higher than the NHC operational model while the false alarm ratio is only 4%, 7% and 6% higher for 25-, 30- and 35-kt RI thresholds, respectively.

References:

Su, H., Wu, L., Jiang, J. H., Pai, R., Liu, A., Zhai, A. J., et al., Applying satellite observations of tropical cyclone internal structures to rapid intensification forecast with machine learning. *Geophysical Research Letters*, 47, e2020GL089102, <http://dx.doi.org/10.1029/2020GL089102> (2020).
Kaplan, J. et al., Evaluating environmental impacts on tropical cyclone rapid intensification predictability utilizing statistical models, *Weather Forecast.* 30 1374–96, <https://doi.org/10.1175/WAF-D-15-0032.1> (2015).

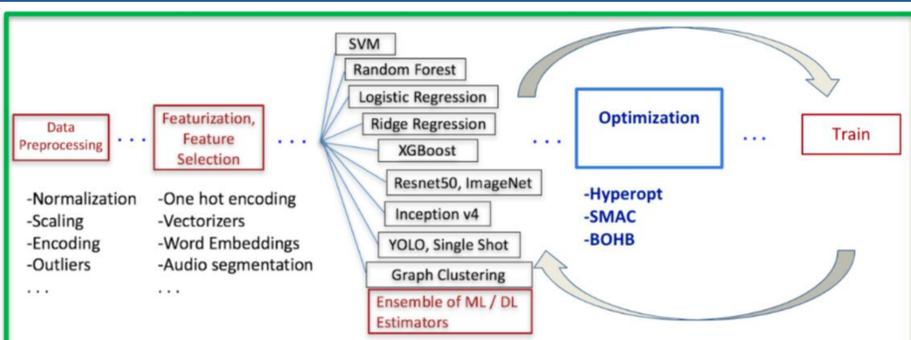
Acknowledgements:

We acknowledge funding support from JPL and in-kind support from IBM and NOAA. The work was conducted at Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA.

Automated Machine Learning (AutoML) as a Service for the Earth Sciences

Technical Lead: **Brian Wilson**

Co-Is: **Alice Yepremyan, Diego Martinez, Sami Sahnoune, Edwin Goh, Sujen Shah, Kai Pak, Santiago Lombeyda, Chris Mattmann, and Wayne Burke**
 Jet Propulsion Laboratory / California Institute of Technology



Assembling an Optimal ML/DL Pipeline

Schematic of AutoML Pipeline Search: preprocessing, featurization/embeddings, feature selection, train an estimator, hyperparameter tuning, and rank by accuracy.

Abstract

As part of the **DARPA D3M program**, JPL is curating a library of ML/DL “primitives” (algorithms) with sufficient metadata and hyperparameter tuning hints to enable auto-assembly (in Python) of pipeline steps. These steps include preprocessing, feature extraction & selection, tuning an ensemble of models, ranking models using a metric, etc. The library contains 90+ classic ML algorithms from scikit-learn, pre-trained deep learning (DL) nets from Keras & PyTorch, and a set of advanced primitives from the D3M performer teams. JPL’s MARVIN tools provide an environment to annotate, discover, install, compose, and execute ML/DL primitives and pipelines. Pipelines and metadata are specified in a declarative manner using a community-defined JSON schema and taxonomy. MARVIN automates the creation of Docker containers containing the primitives and software dependencies, which are executed on a Kubernetes cluster either on premise or at any Cloud vendor supporting Kubernetes. D3M is designed to solve 15+ problem types:

- Classification, Regression, Clustering
- Image classification, object recognition
- Graph clustering/matching, Recommendations, Links
- Audio segmentation, video processing
- Time-series forecasting etc.

Exploring the library of ML algorithms, datasets/problems, pipelines

MARVIN enables an Automated ML environment similar to an “app store” in which a new “discoverable” ML/DL capability can be added by authoring a simple Python class satisfying the method interface, with tuning hints and a bit of metadata from the taxonomy. Currently, MARVIN contains 600+ datasets/problems, 330+ primitives, and 5 Million+ Pipeline Runs.

MARVIN

An Open Machine Learning Corp. Primitive Annotation and Execution Framework

Explore our metalearning database:

Datasets - 5165 results

Problems - 3939 results

Primitives - 2877 results

Pipeline Runs - 5362052 results

Other resources:

Docker Registry

Metalearning Information

The screenshots show the MARVIN web interface. The top section displays search filters for version (4.0.0) and metrics (F1_MACRO). Below this, there are three main sections: **Datasets** (showing 'World development indicators: Life expectancy prediction dataset'), **Problems** (showing 'timeseries classification problem' and 'LL1 Cinc ECG torso problem'), and **Primitives** (showing 'sklearn.neural_network.multilayer_perceptron.MLPClassifier' and 'BERT pair classification'). Each section includes search bars, filters, and lists of results with links to details.

5M+ Pipeline Runs

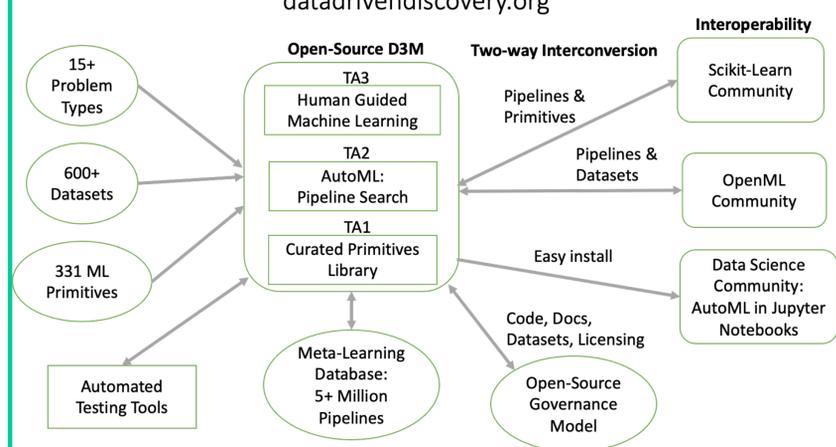
- Ranked by score
- Cmp pipelines
- Cmp key primitives
- Cmp hyperparameters
- Group by problem type & team
- Leaderboards

Future Work

- Inject Earth science remote sensing problems -- “phenomena recognition”, anomaly detection and time-series forecasting problems -- into the D3M program.
- Soliciting datasets and problems.
- Enable Meta-learning across 10M+ solution pipelines to be used for future model selection.
- GUI’s for domain experts: einblick.ai, Harvard Two Ravens, AutoML from Jupyter Notebooks

D3M Ecosystem and Community Outreach

datadrivendiscovery.org



Poster 13: Development of Gap-agnostic Machine Learning techniques

for Earth Science applications

Soni Yatheendradas^{1,2}, Sujay Kumar², Christa Peters-lidard², and David Mocko^{2,3}

¹UMD ESSIC, ²NASA GSFC, ³SAIC



1. Introduction

- Input data gaps hamper machine learning
- Current gap-agnostic technique of partial convolution can handle pixels where all inputs have valid values
- But not effective for earth sciences where data gaps vary across inputs!
- For full gap-agnosticity, we developed generalized versions of partial convolution, partial neural network layer, and input relevance calculation
- Our example applications:
 - Regression: Gap-filling remotely-sensed snow cover fraction (SCF) through downscaling
 - Classification: Using Climate Prediction Center (CPC) Experimental Objective Blend inputs for predicting United States Drought Monitor (USDM) categories

2. Data and methodology

A. SCF gap-filling using Super-Resolution Convolutional Neural Network

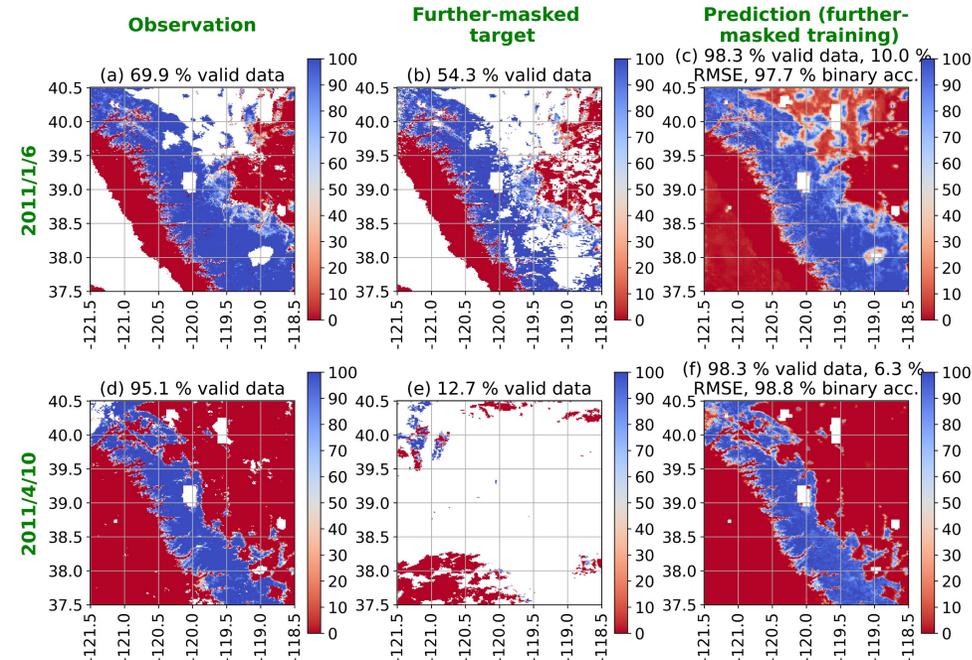
- Target: MOD10A1 C5 SCF (1-km version)
- Core input/s: 5-km MOD10A1 C5 SCF; SCF, Cloud Cover Fraction (CCF), Confidence Index (CI)
- Auxiliary inputs (1 km resolution):
 - static terrain: elevation, slope, aspect
 - Land Surface Model dynamics (Noah-MP): Precip, Snow Water Equivalent (SWE), surface radiative temperature, Leaf Area Index (LAI)
 - Satellite-based: C5: MOD10A1 snow albedo, MOD11A1 land surface temperature (LST)
- 3° X 3° domain centered over Lake Tahoe (CA-NV border), 3- year training (2009-2011), 2012 for validation, 2005-2007 for further synthetic cloud masking

B. Input relevances in USDM classification

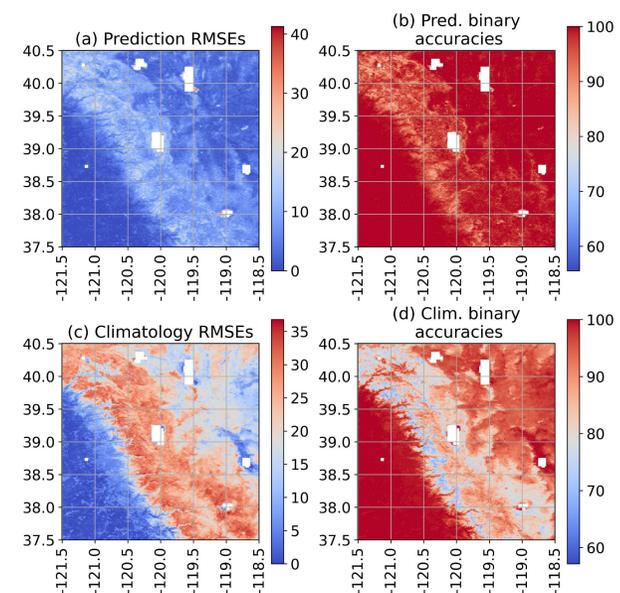
- Data resolved to US Climate Division polygons
- CPC Blends' Short- and Long-Term inputs: Z-index; 60-month Z-index; Modified and Hydrologic Palmer Drought Indices (PMDI, PHDI); Precipitations for 1, 3, 6, 12, 24 and 60 months; CPC soil moisture
- Training years 2006-2018, Validation 2019
- 1 intermediate layer of 16 neurons
- Our modified Softmax-Gradient Layer-wise Relevance Propagation (SGLRP) for better input contrast and relevance conservation

3a. Results (SCF gap-filling)

• Example-day predictions



• Spatial error: ML-based prediction vs. climatology



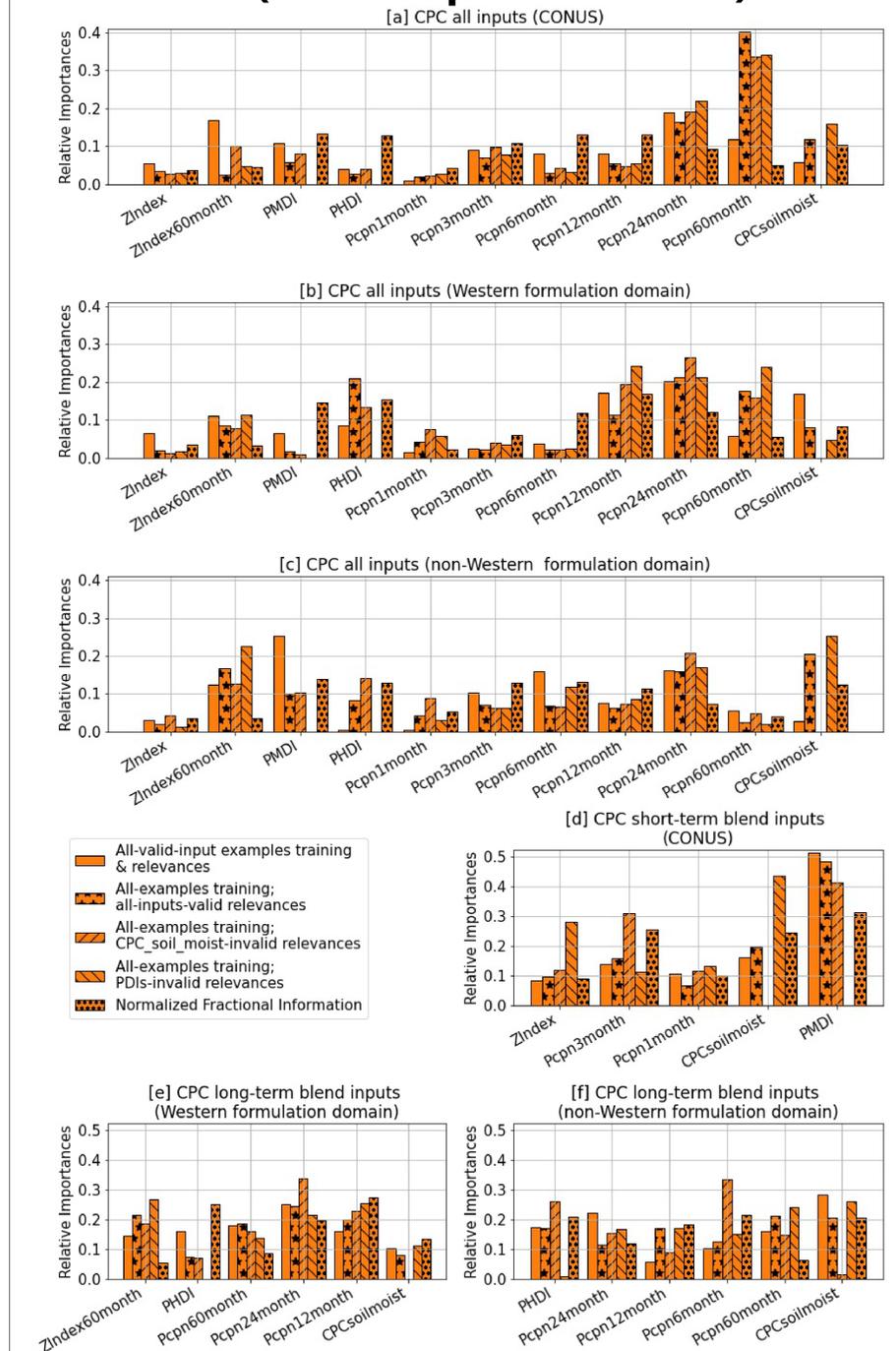
4. Summary

- SCF gap-filling:
 - Our generalized technique successfully recreates synthetically masked-out areas much better than the climatology baseline
- USDM input relevances:
 - Regular training bars (no-hatch) and fully gap-agnostic ones (star and diagonal hatchings) show different information
- These advancements provide the earth science community with a gap-agnostic Machine Learning infrastructure tool

5. References

- Iwana, B. K. et al (2019), Explaining Convolutional Neural Networks using Softmax Gradient Layer-wise Relevance Propagation, arXiv:1908.04351
- Liu, G., et al. (2018), Image Inpainting for Irregular Holes Using Partial Convolutions, arXiv:1804.07723
- Vandal, T., et al. (2017), DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution, DOI: <http://dx.doi.org/10.1145/3097983.3098004>

3b. Results (USDM input relevances)



Motivation

Identification of flood water extent from satellite images has historically relied on either synthetic aperture radar (SAR) or multi-spectral (MS) imagery. But MS sensors may not penetrate cloud cover, whereas SAR is plagued by operational errors such as noise-like speckle challenging their viability to global flood mapping applications. An attractive alternative is to effectively combine MS data and SAR, i.e., two aspects that can be considered complementary with respect to flood mapping tasks. Therefore, in this study, we explore the diverse bands of Sentinel 2 (S2) derived water indices and Sentinel 1 (S1) derived SAR imagery along with their combinations to access their capability in generating accurate flood inundation maps using a fully connected deep convolutional neural network known as U-Net.

Data

For this study, we use a new georeferenced flood label data i.e. Sen1Floods11 (Bonafieil et al., 2020) which provides flood inundation labels spanning over 11 flood events across the world (Table 1, Figure 1). This dataset contains human supervised flood labels generated for 446 images at 10-meter resolution at 512x512 dimensions. The dataset further provides corresponding S1 and S2 bands for the labeled flood events. For our study, we only use all the 446 human supervised images for deep learning model evaluation.

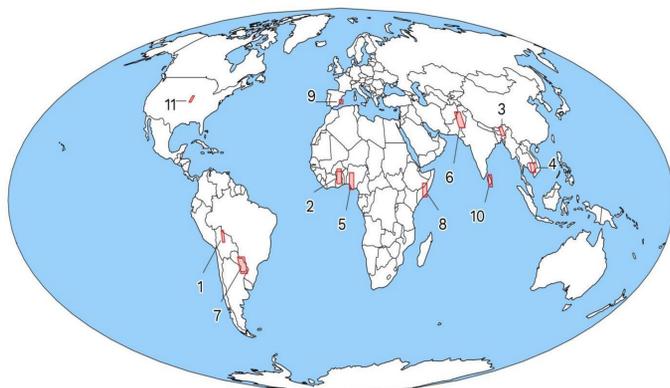


Figure 1: Locations of flood events sampled in Sen1Floods11

Table 1: Flood event acquisition time in Sen1Floods11

ID	Country	S2 Date	S1 Date
1	BOLIVIA	2/15/2018	2/15/2018
2	GHANA	9/19/2018	9/18/2018
3	INDIA	8/12/2016	8/12/2016
4	VIETNAM	8/4/2018	8/5/2018
5	NIGERIA	9/20/2018	9/21/2018
6	PAKISTAN	6/28/2017	6/28/2017
7	PARAGUAY	10/31/2018	10/31/2018
8	SOMALIA	5/5/2018	5/7/2018
9	SPAIN	9/18/2019	9/17/2019
10	SRI LANKA	5/28/2017	5/30/2017
11	USA	5/22/2019	5/22/2019

U-Net

We use a variation of Convolutional neural network architectures called as U-Net for flood identification. U-Net has been referred to as having an effective structure to successfully perform image segmentation tasks (Ronneberger et al., 2016). Firstly, the encoder half of the model carries out a downsampling process, bringing the input image down to small size feature matrix (Figure 2). Secondly, the decoder constructs the model output using the features as input and carries out an upsampling process to bring back the spatial information of input image

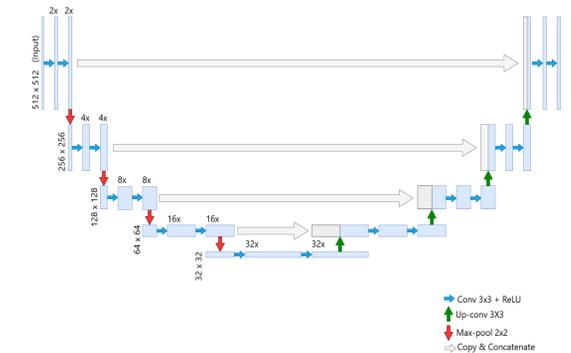


Figure 2: Pictorial representation of U-net architecture adopted for segmentation of water extents

Experimental setup & evaluation criteria

The 446 images used in this study are divided randomly into splits containing 70 % of images for training, 20 % for validating and 10 % for testing. Several possible combinations were tested. For all the combinations, a separate set of experiments with DEM as an additional input to combination bands of S-1 and S-2 were also run. Firstly, in case of S1, we use both VV/VH bands to establish benchmark performance of S-1 for flood inundation mapping. In case of S-2, the combined spectral indices of NDVI and MNDVI (cNDVI), $AWEI_{sh}$ and $AWEI_{nsh}$ (cAWEI) and their combination as input to our deep learning algorithm. Also, HSV bands and their combination with spectral indices were used in our experiment. Finally, all S-2 band configurations are combined with S-1 to evaluate S-1 and S-2 combinations. Precision, recall and their harmonic mean (F1 Score) were used to evaluate combinations (Figure 3).

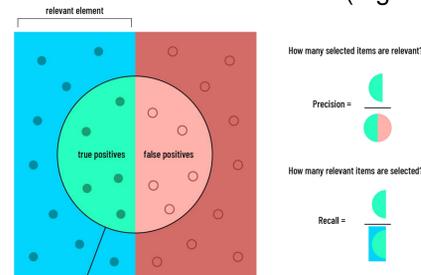


Figure 3: Schematic showing calculation of evaluation metrics

Results

Impact of DEM on S1

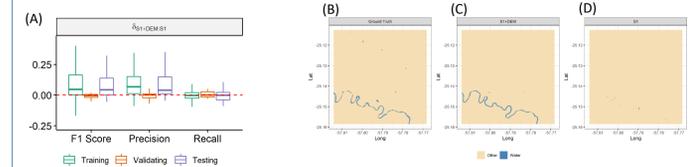


Figure 4: The fractional difference between performance of S1 bands+DEM and S1 is shown in (A). Flood inundation of a location in Paraguay based on ground truth (B), produced by U-Net when using (C) S1+DEM as input and (D) HSV + DEM as input.

Difference in performance between S2 and S1

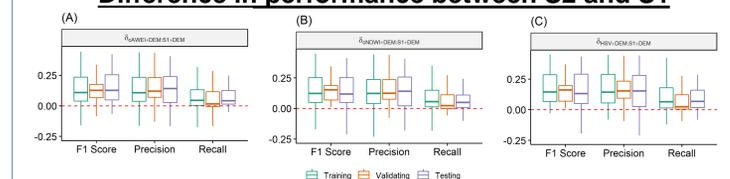


Figure 5: Boxplots representing the fractional difference between (A) cAWEI+DEM (B) cNDVI+DEM (C) HSV+DEM and S1+DEM.

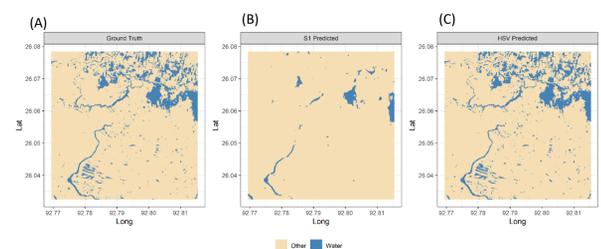


Figure 6: Flood inundation of a location in India based on ground truth (A), produced by U-Net when using (B) S1+DEM as input and (C) HSV + DEM as input.

Performance change between S1+S2 and S2

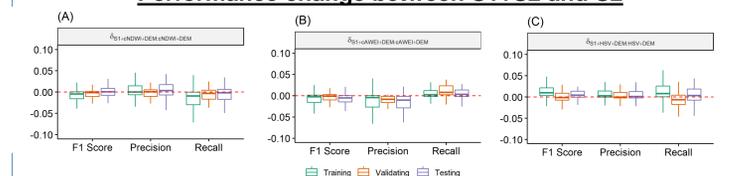


Figure 7: Boxplots representing the fractional difference in performance between (A) S1+cNDVI+DEM and cNDVI+DEM, (B) S1+cAWEI+DEM and cAWEI+DEM (C) S1+HSV+DEM and HSV+DEM

Performance improvement of between S1+S2 over MODIS NRT

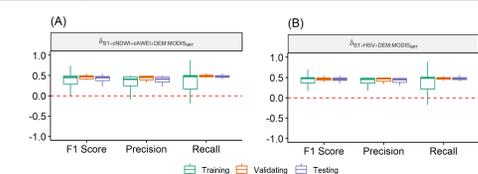


Figure 8: Boxplots representing the fractional difference in performance between (A) S1+cAWEI+cNDVI+DEM (B) S1+HSV+DEM and MODIS NRT

Summary:

- Our results indicated that even though the SAR data is not affected by cloud cover, poor contrast between VV and VH backscatter has affected S1 data's flood inundation mapping performance.
- The trained U-net was able to achieve a median F1 score of 0.74 when using DEM and S1 bands as input in comparison to 0.63 when using only S1 bands highlighting the active positive role of DEM in mapping floods.
- Among the S2 bands, HSV (Hue, Saturation, Value) transformation of Sentinel 2 data has achieved a median F1 score of 0.91 outperforming the commonly used water spectral indices owing to HSV's transformation's superior contrast distinguishing abilities.
- Also, the U-Net algorithm outperforms the MODIS NRT products by around 50%.

References:

- Bonafilia, D., Tellman, B., Anderson, T., & Issenberg, E. (2020). Sen1Floods11: a georeferenced dataset to train and test deep learning flood algorithms for Sentinel-1. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 835-845.
- O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, pp. 234-241, 2015.
- Konapala G., Kumar S.V., Ahmed S., Exploring Sentinel-1 and Sentinel-2 diversity for Flood inundation mapping using deep learning, (In review)