

Accelerating Earth and climate modeling with machine learning

Kelly Kochanski

NCAR Multicore Workshop 2019

WHEN A USER TAKES A PHOTO,
THE APP SHOULD CHECK WHETHER
THEY'RE IN A NATIONAL PARK...

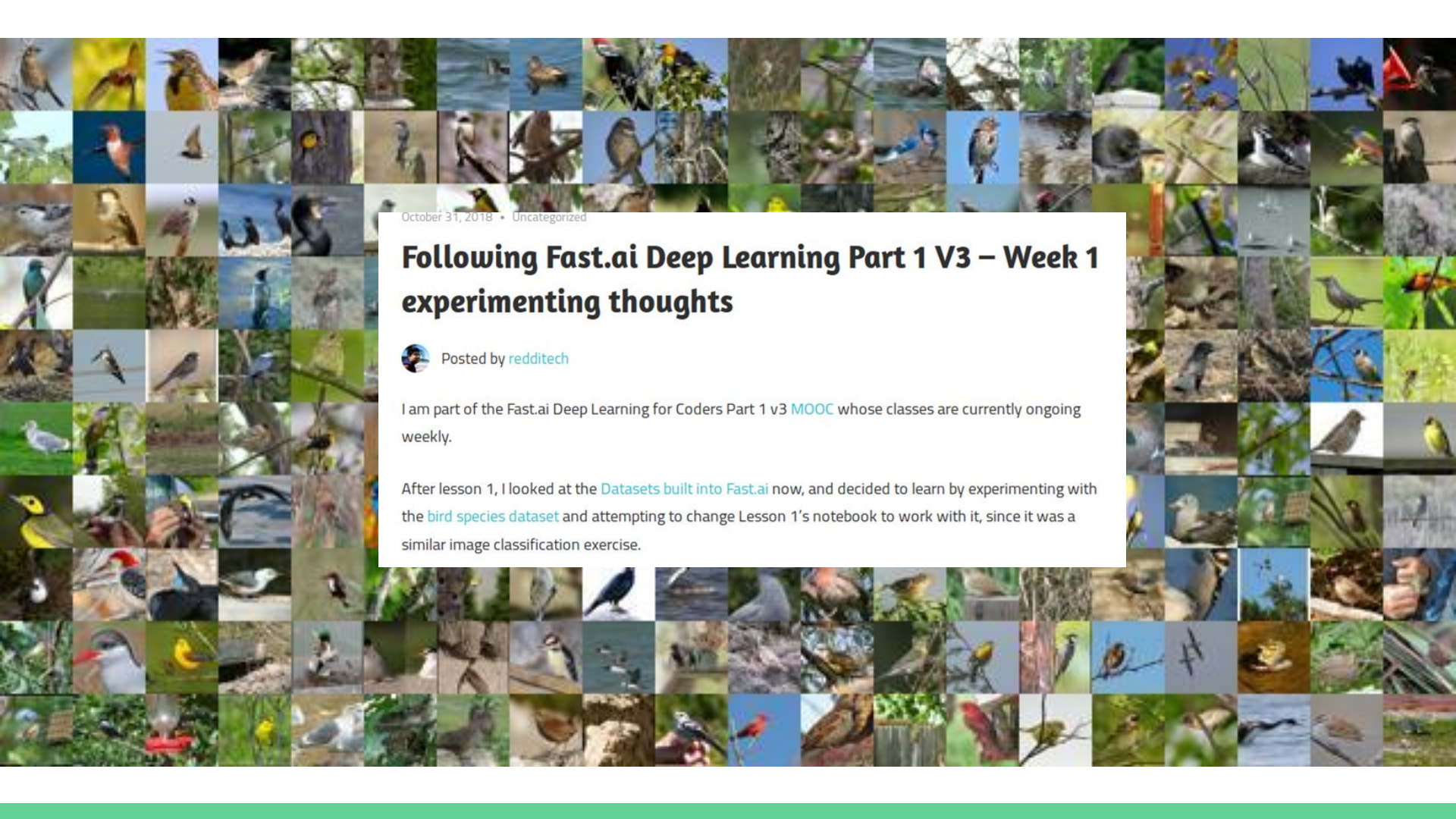
SURE, EASY GIS LOOKUP.
GIMME A FEW HOURS.

... AND CHECK WHETHER
THE PHOTO IS OF A BIRD.

I'LL NEED A RESEARCH
TEAM AND FIVE YEARS.



IN CS, IT CAN BE HARD TO EXPLAIN
THE DIFFERENCE BETWEEN THE EASY
AND THE VIRTUALLY IMPOSSIBLE.



October 31, 2018 • Uncategorized

Following Fast.ai Deep Learning Part 1 V3 – Week 1 experimenting thoughts

 Posted by [redditech](#)

I am part of the Fast.ai Deep Learning for Coders Part 1 v3 [MOOC](#) whose classes are currently ongoing weekly.

After lesson 1, I looked at the [Datasets built into Fast.ai](#) now, and decided to learn by experimenting with the [bird species dataset](#) and attempting to change Lesson 1's notebook to work with it, since it was a similar image classification exercise.



What is machine learning?



What is machine learning?

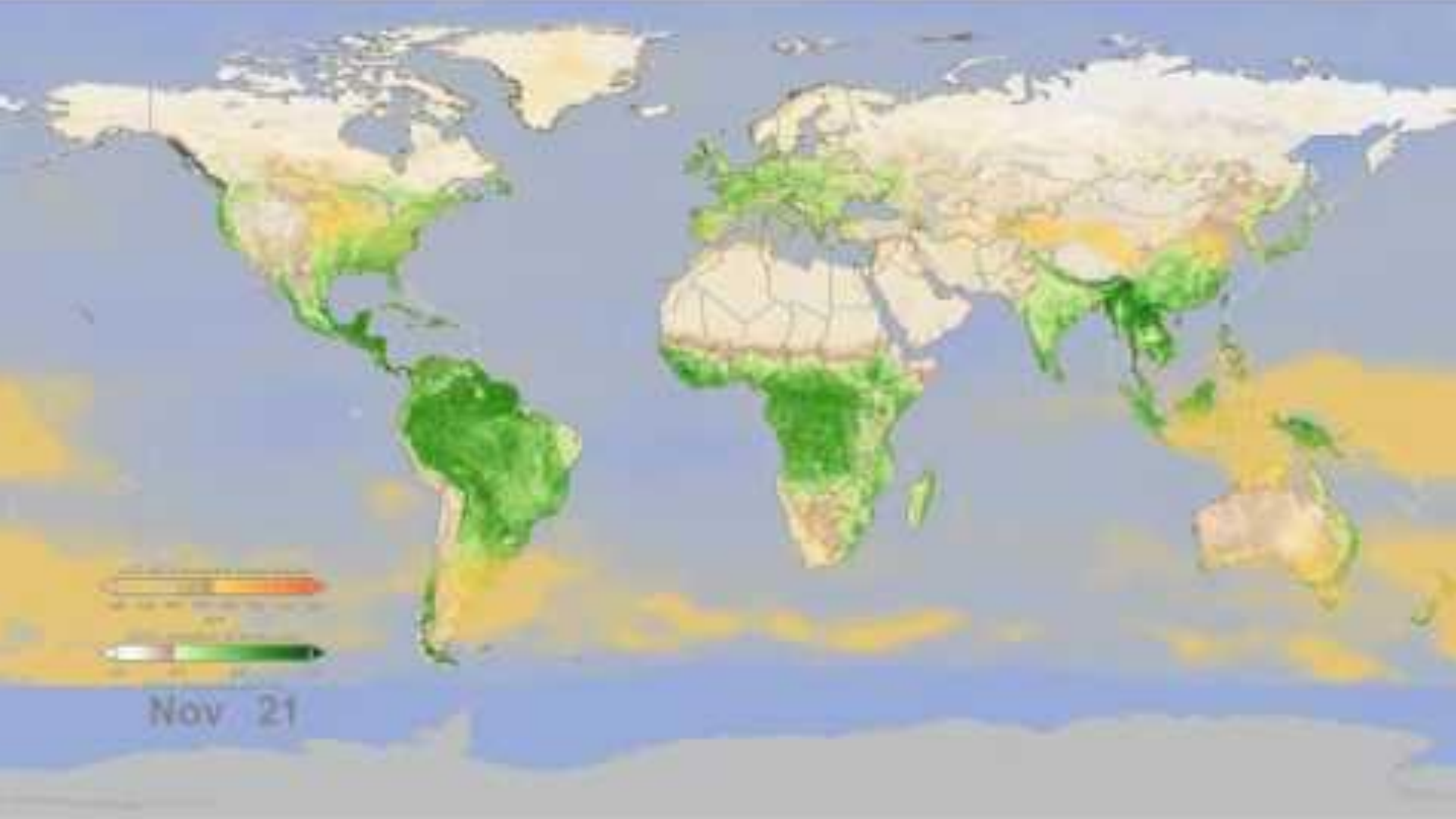
Machine learning at its most basic is the practice of using algorithms to parse data, learn from it, and then make a determination or prediction about something in the world.

Why is machine learning relevant
to Earth System Modeling now?



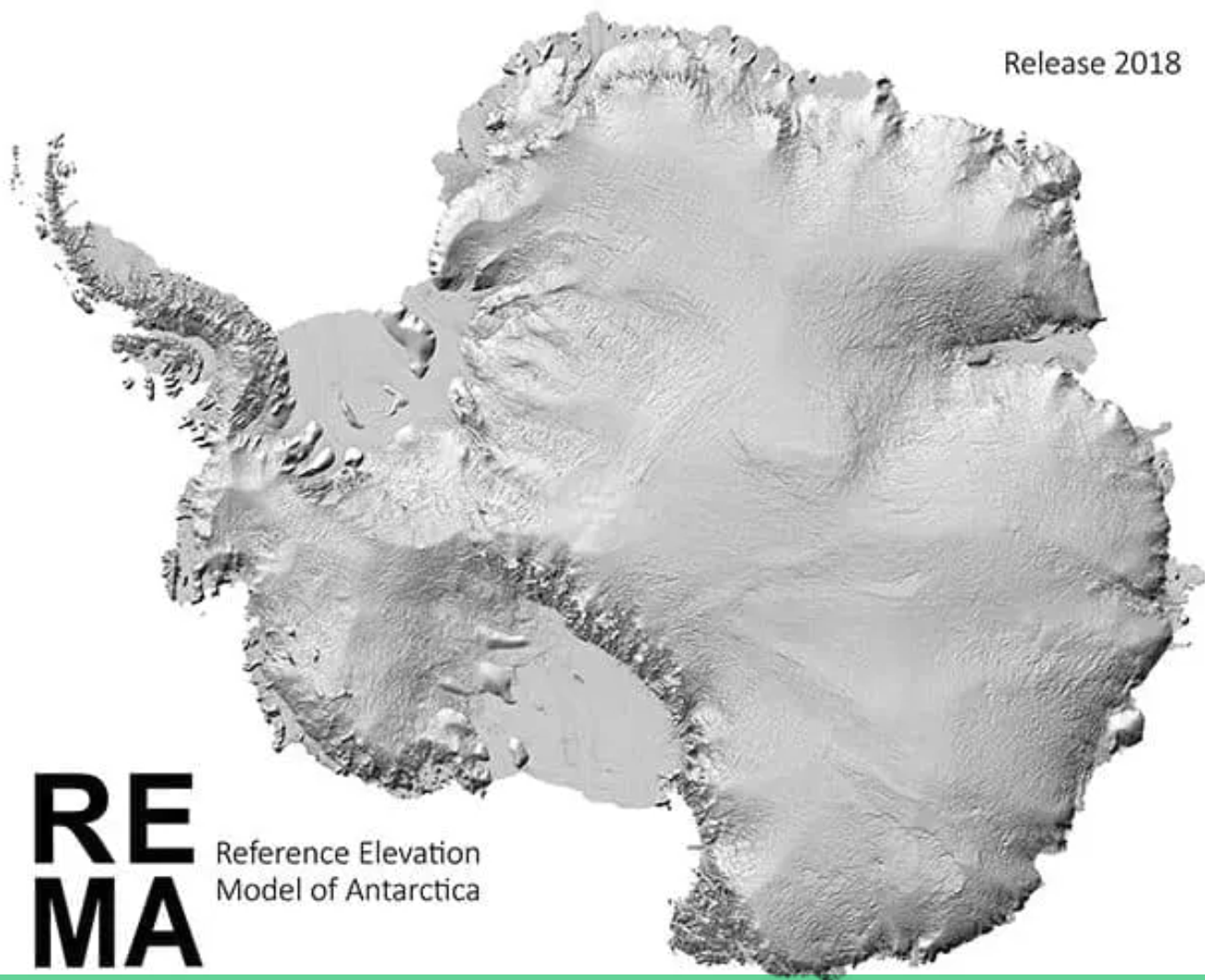
Current trends 1/3

Machine learning
offers solutions to
once-intractable
problems



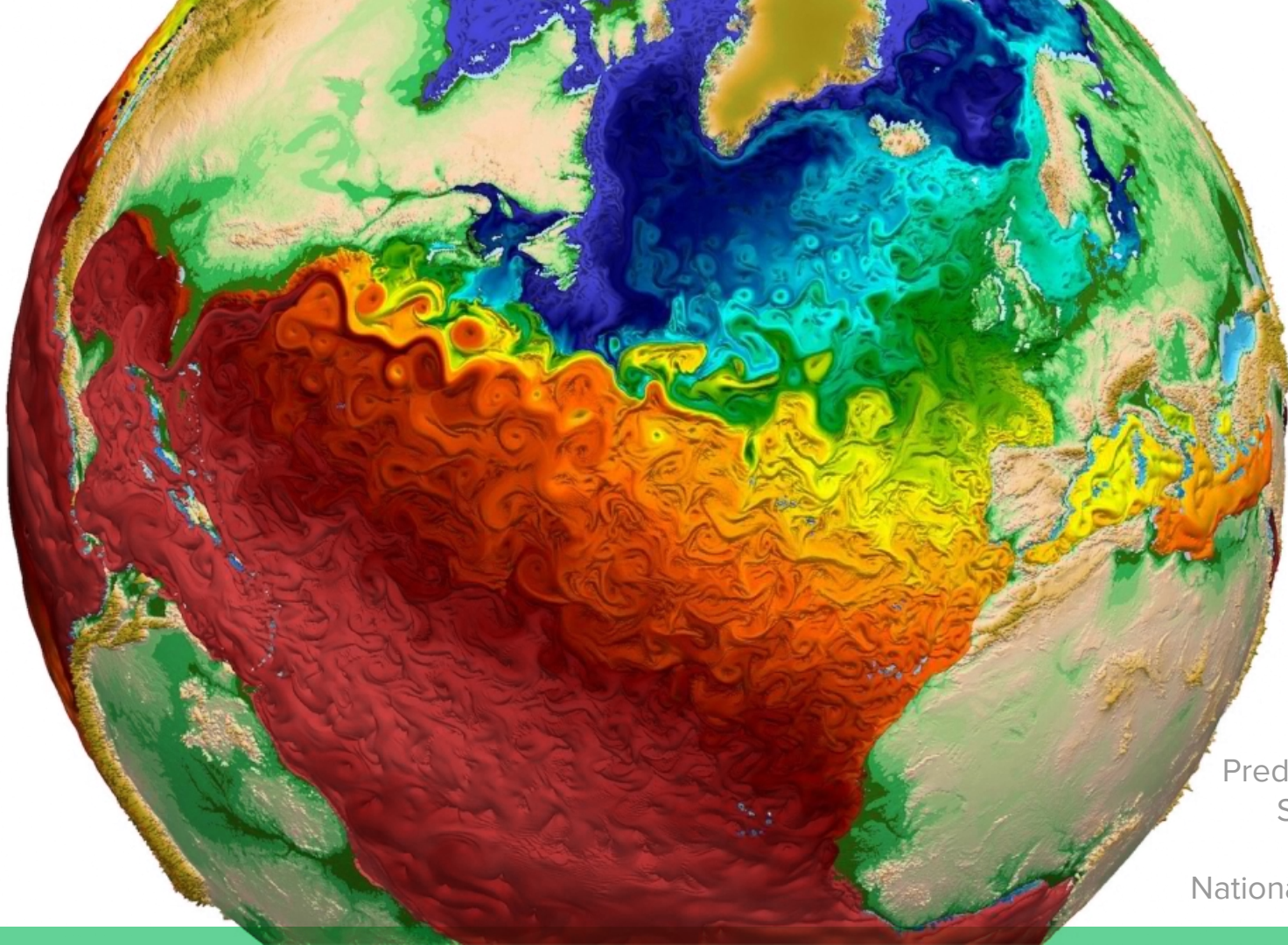
Nov 21

Release 2018



**RE
MA**

Reference Elevation
Model of Antarctica

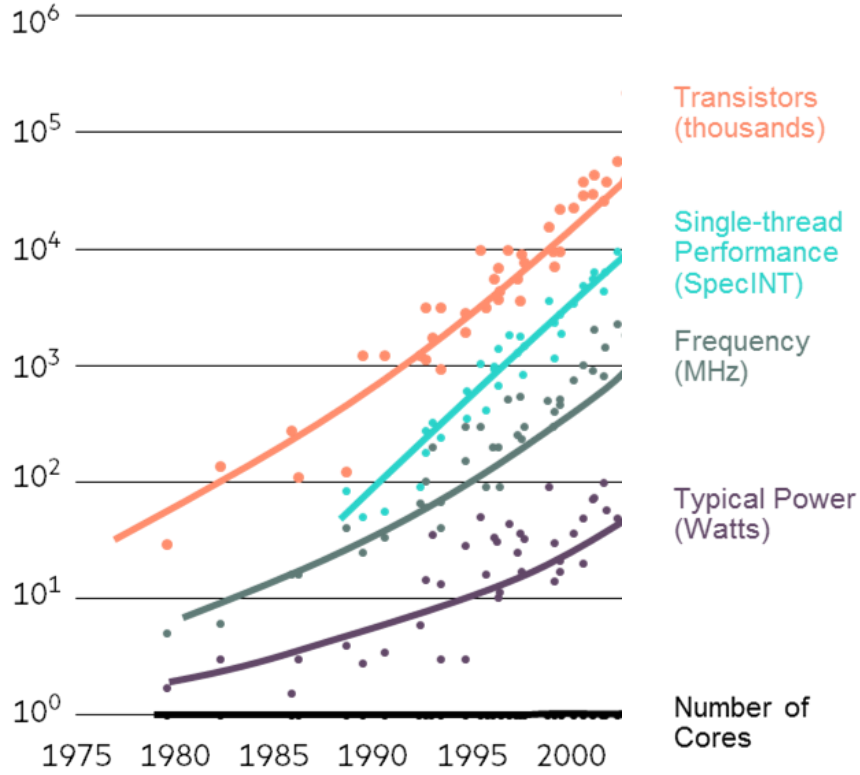


Model for
Prediction Across
Scales (2015),
Los Alamos
National Laboratory

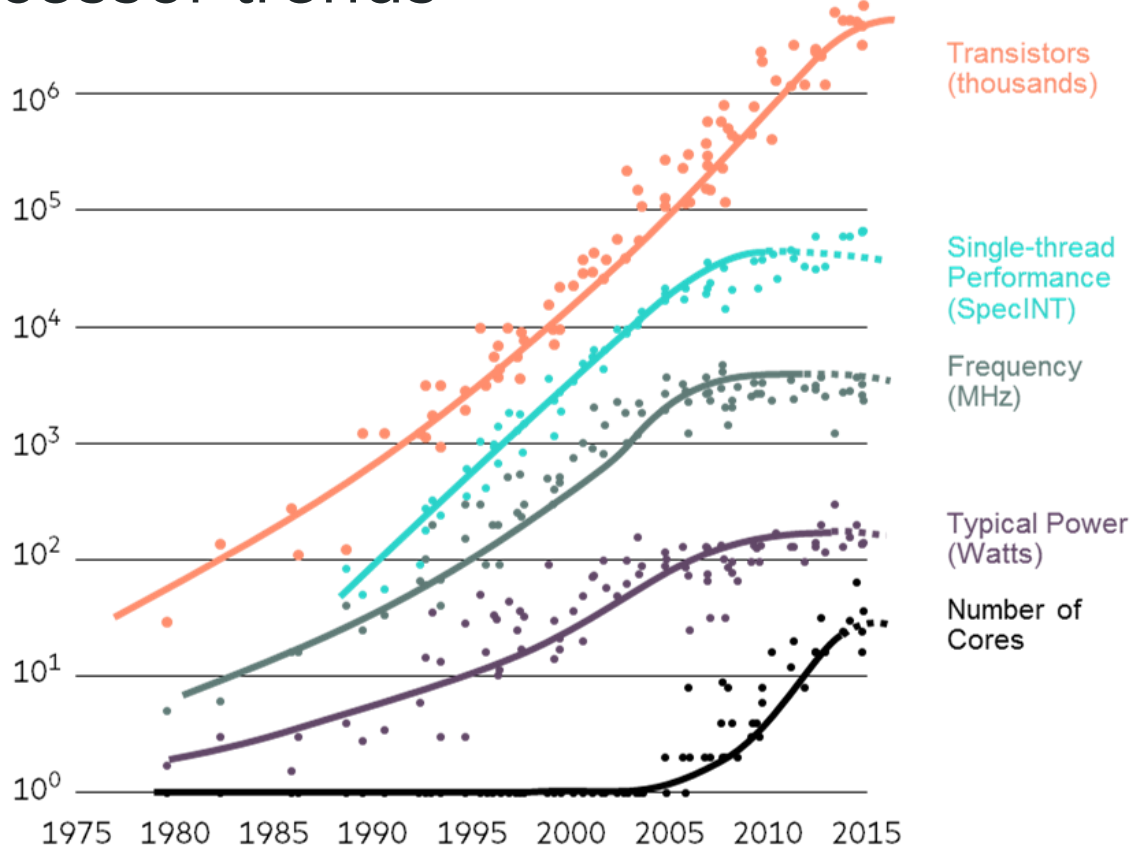
Current trends 2/3

New data streams
increase the potential
power of data-driven
models

Microprocessor trends



Microprocessor trends



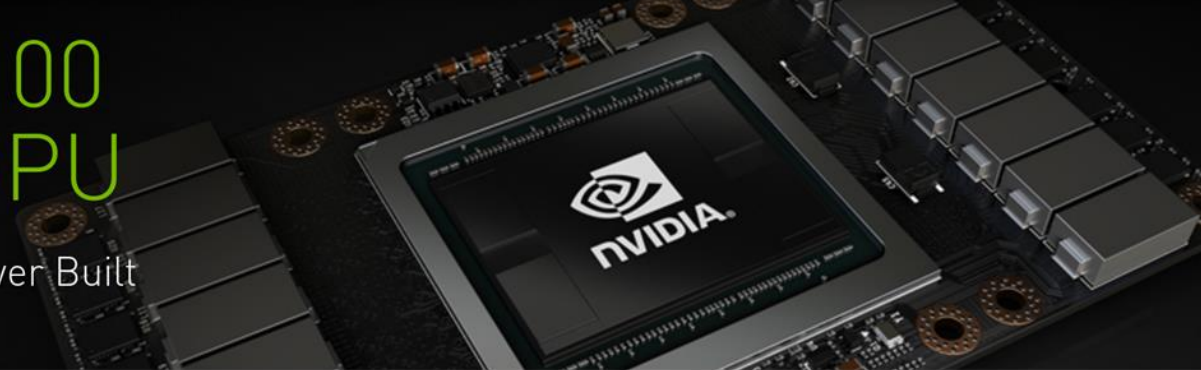


9,216 Power9 22-core CPUs

27,648 NVIDIA Tesla V100
GPUs

NVIDIA TESLA V100 TENSOR CORE GPU

The Most Advanced Data Center GPU Ever Built



WELCOME TO THE ERA OF AI.

Finding the insights hidden in oceans of data can transform entire industries, from personalized cancer therapy to helping virtual personal assistants converse naturally and predicting the next big hurricane.

Google TensorFlow Processing Units



IBM TrueNorth Chips



Current trends 3/3

Machine learning is
driving innovation in
HPC

My perspective:
Climate change impacts
ML in service of earth science



Tackling Climate Change with Machine Learning

David Rolnick^{1*}, Priya L. Donti², Lynn H. Kaack³, Kelly Kochanski⁴, Alexandre Lacoste⁵,
Kris Sankaran^{6,7}, Andrew Slavin Ross⁸, Nikola Milojevic-Dupont^{9,10}, Natasha Jaques¹¹,
Anna Waldman-Brown¹¹, Alexandra Luccioni^{6,7}, Tegan Maharaj^{6,7}, Evan D. Sherwin²,
S. Karthik Mukkavilli^{6,7}, Konrad P. Kording¹, Carla Gomes¹², Andrew Y. Ng¹³,
Demis Hassabis¹⁴, John C. Platt¹⁵, Felix Creutzig^{9,10}, Jennifer Chayes¹⁶, Yoshua Bengio^{6,7}

¹University of Pennsylvania, ²Carnegie Mellon University, ³ETH Zürich, ⁴University of Colorado Boulder,

⁵Element AI, ⁶Mila, ⁷Université de Montréal, ⁸Harvard University,

⁹Mercator Research Institute on Global Commons and Climate Change, ¹⁰Technische Universität Berlin,

¹¹Massachusetts Institute of Technology, ¹²Cornell University, ¹³Stanford University,

¹⁴DeepMind, ¹⁵Google AI, ¹⁶Microsoft Research

climatechange.ai

How can we use machine learning to build better Earth System Models?

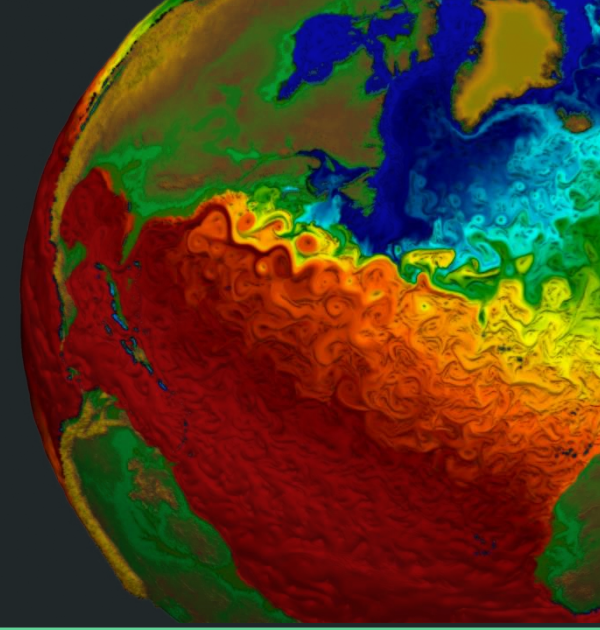
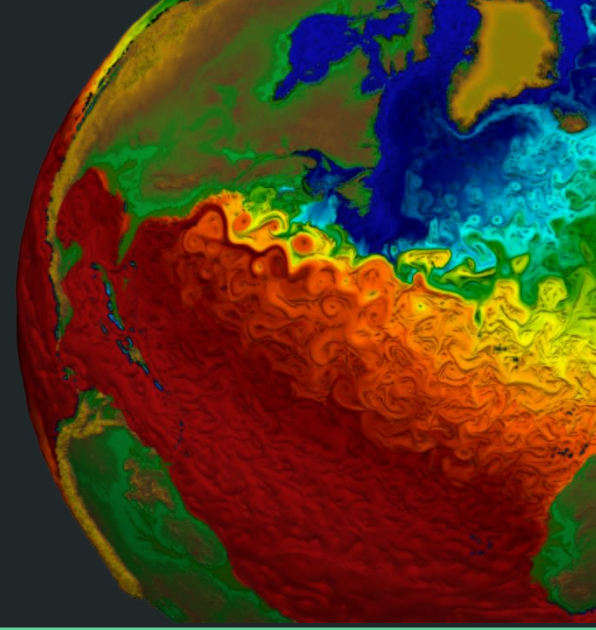


Image: MPAS-Ocean
Los Alamos National Lab

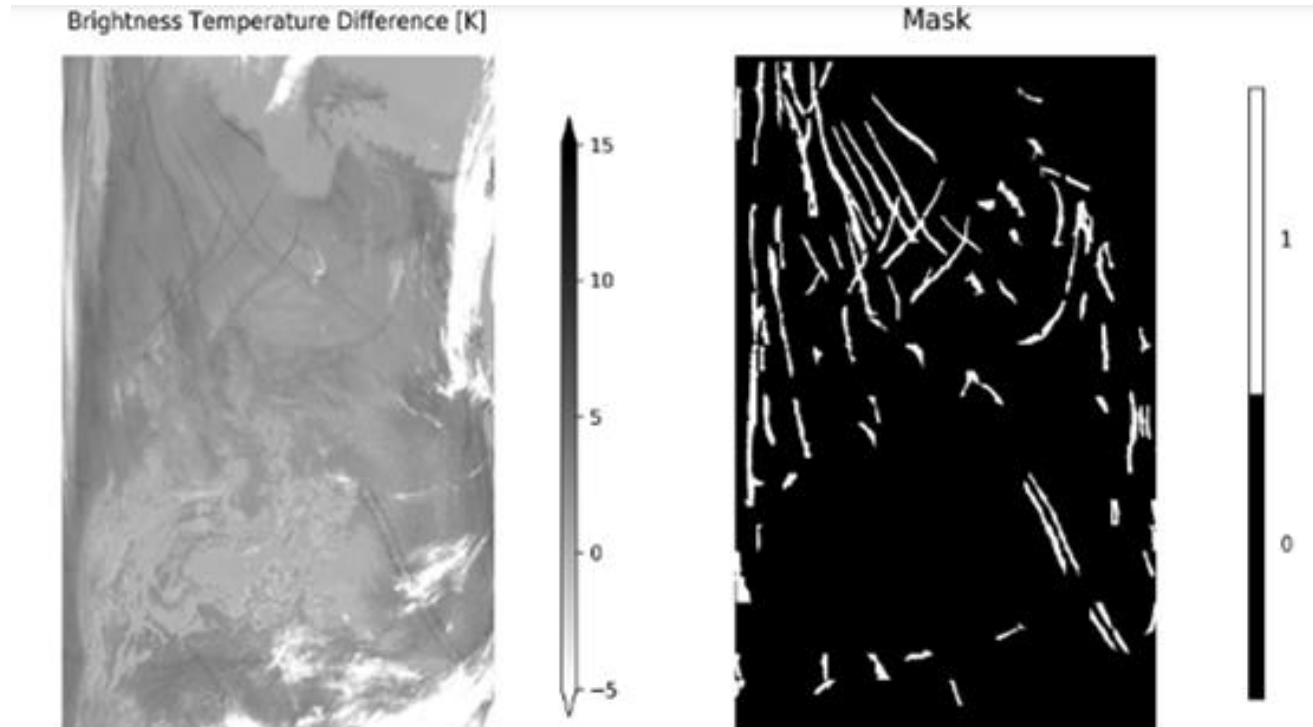
How can we use machine learning to build better Earth System Models?



Aims:

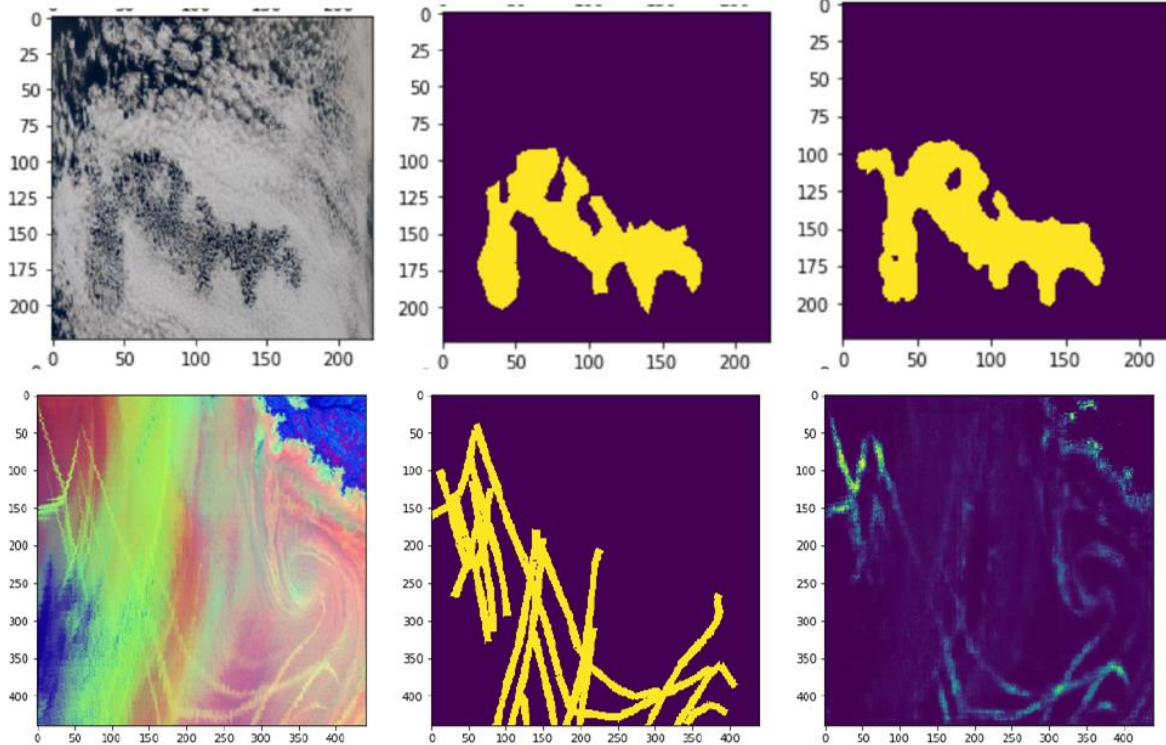
- To solve long-standing problems with new methods
- To integrate new sources of data into existing models
- To take advantage of new computing hardware

Monitoring marine clouds



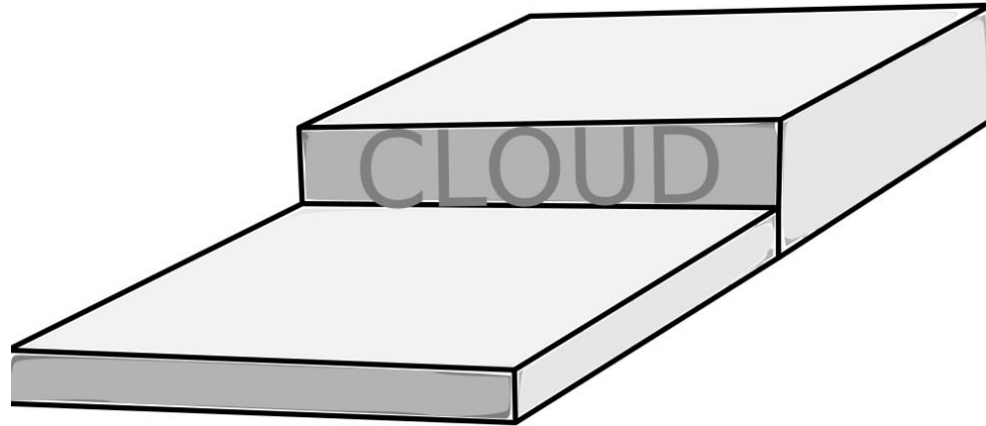
Yuan, Tianle, et al.
"Automatically Finding
Ship-tracks to Enable
Large-scale Analysis of
Aerosol-Cloud Interactions."
*Geophysical Research
Letters* (2019).

Monitoring marine clouds

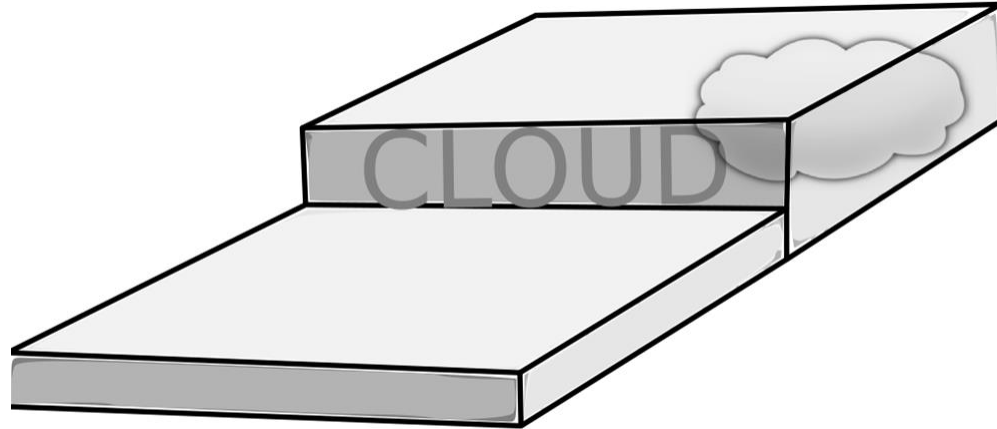


Watson-Parris, Duncan, et al.
"Detecting anthropogenic
cloud perturbations
with deep learning."
International Conference on
Machine Learning, 2019.

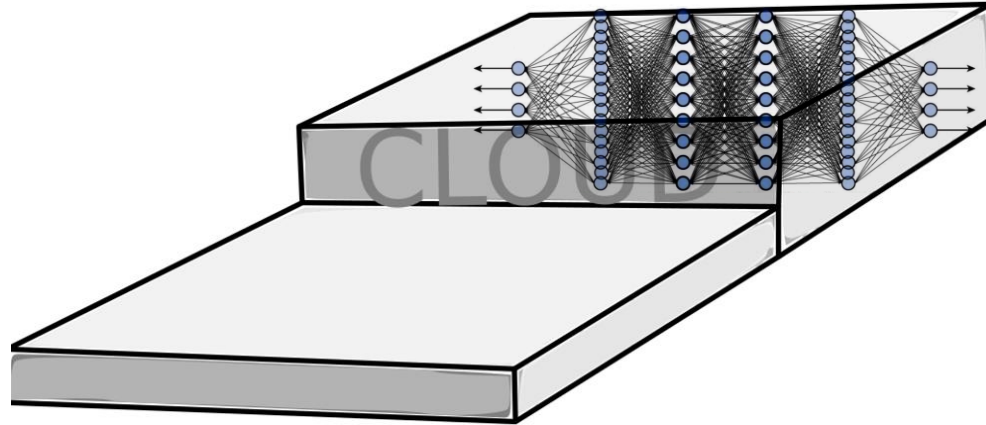
Improving convection + aerosol modelling



Improving convection + aerosol modelling

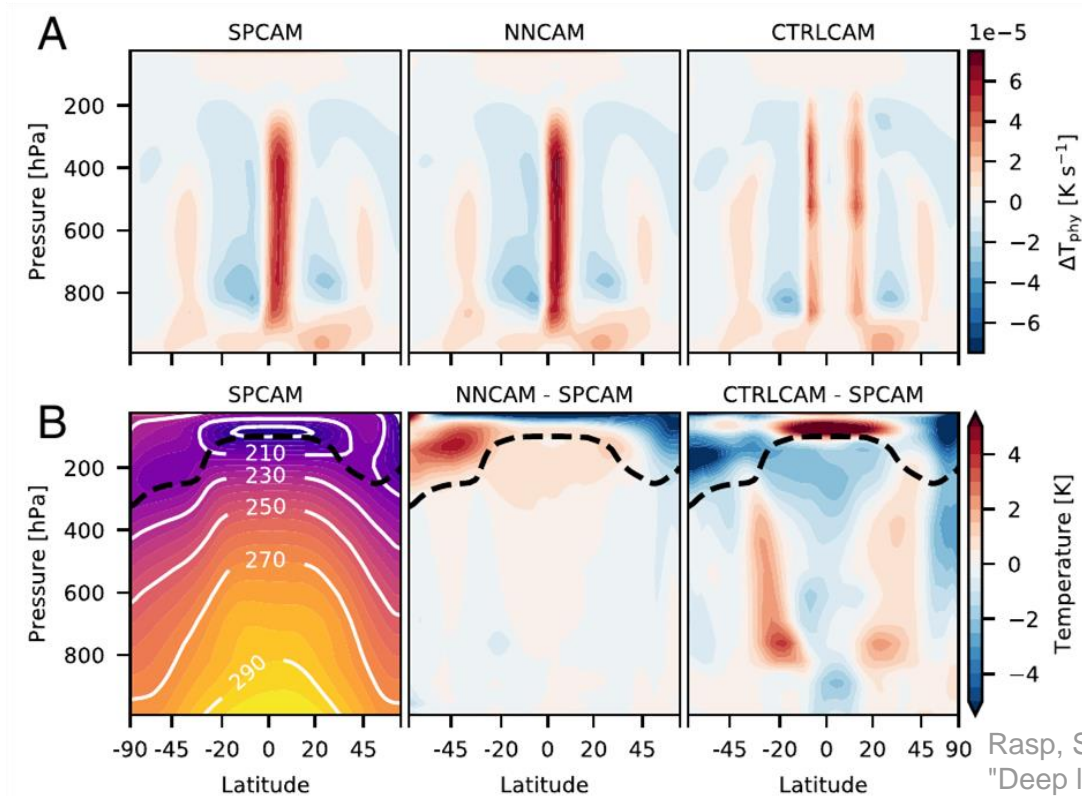


Improving convection + aerosol modelling



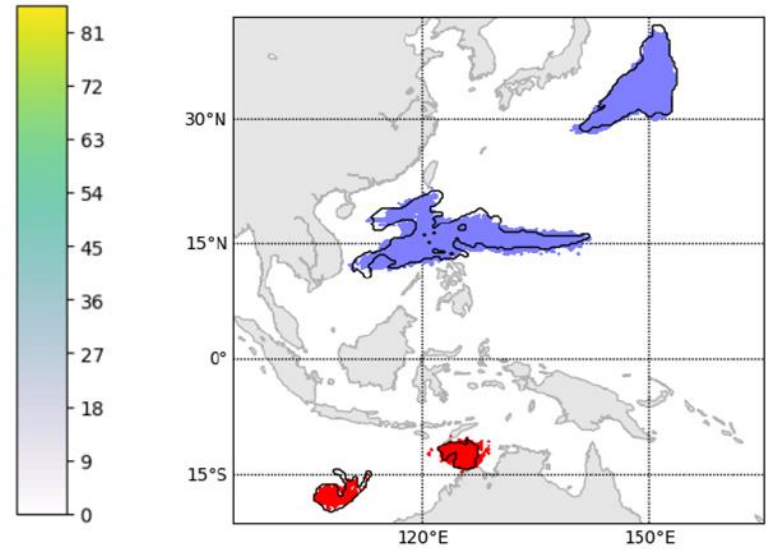
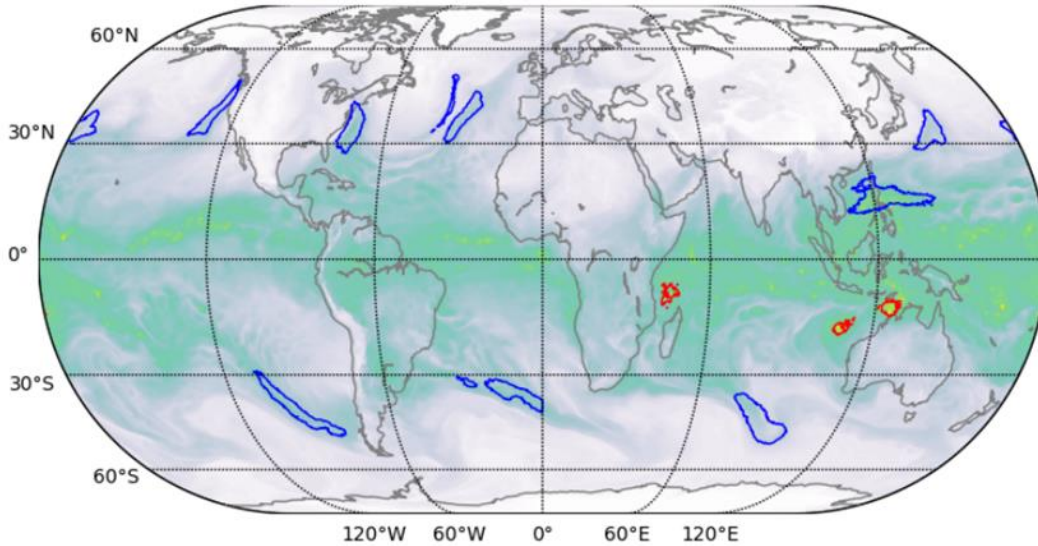
Gentine, Pierre, et al. "Could machine learning break the convection parameterization deadlock?." *Geophysical Research Letters* 45.11 (2018): 5742-5751.

Improving convection + aerosol modelling



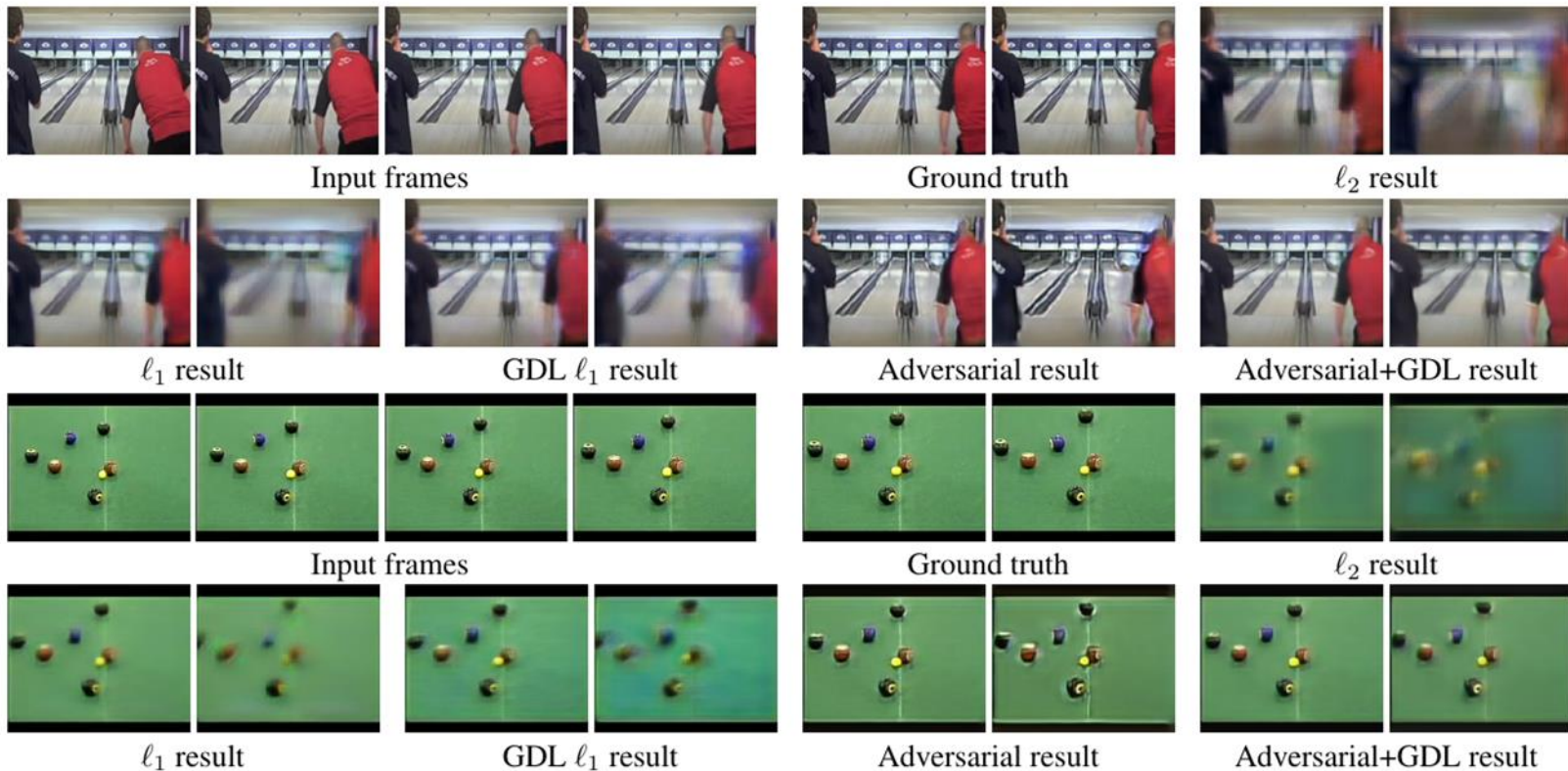
Rasp, S, M. S. Pritchard, and P. Gentine.
"Deep learning to represent subgrid
processes in climate models." *PNAS* (2018)

Tracking extreme events



Kurth, Thorsten, et al. "Exascale deep learning for climate analytics." *Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis*. IEEE Press, 2018.

Deep learning for spatio-temporal patterns

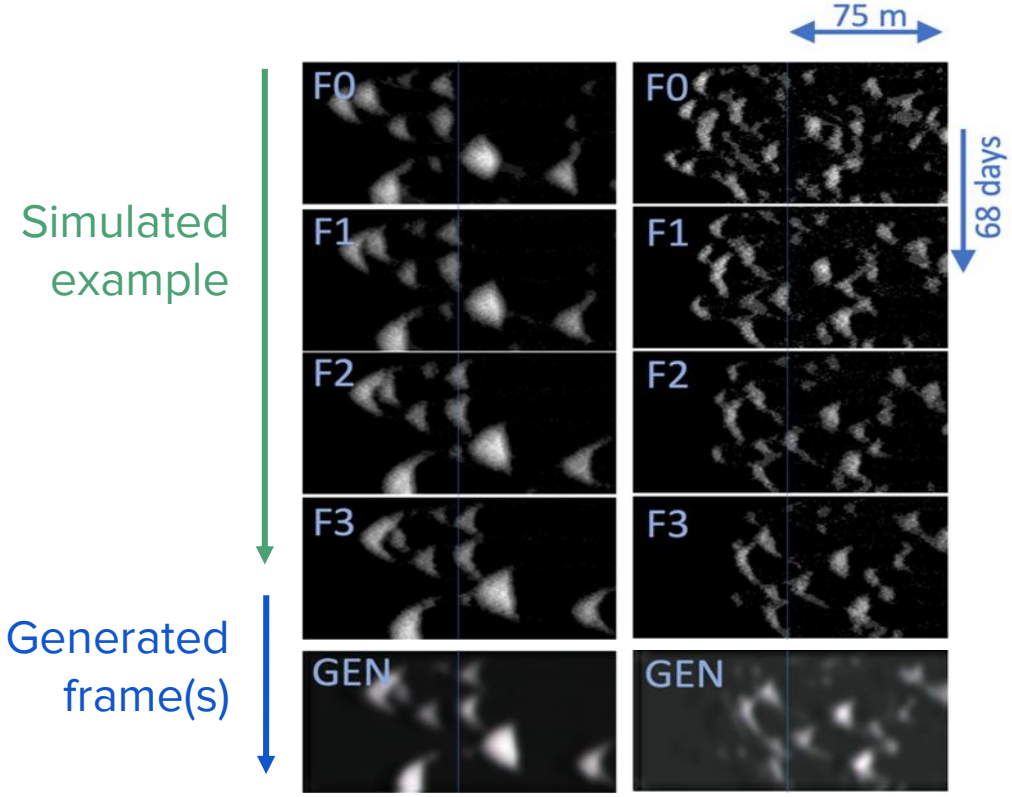


Deep learning for spatio-temporal patterns



Reedster, Mogle and Bogel, 'Monitoring and analysis of sand dune movement and growth on the Navajo Nation, Southwestern United States' (2011) USGS Fact Sheet 3085.

Deep learning for spatio-temporal patterns



Barriers to implementation

Barriers to implementation

Machine learning

Climate science

What's exciting?

Big data!

Science!

Objectives

Well-defined is useful.

Broad is interesting.

Explainability

Second to prediction

Often the main goal

Data

Ideally clean and labelled

Many unlabeled features

Data formats

Images, csv, dataframes

Images, netcdf, misc

Data use

Integral to model

Data -> theory -> model

Existing code

Python, R, Julia

C/C++, Fortran

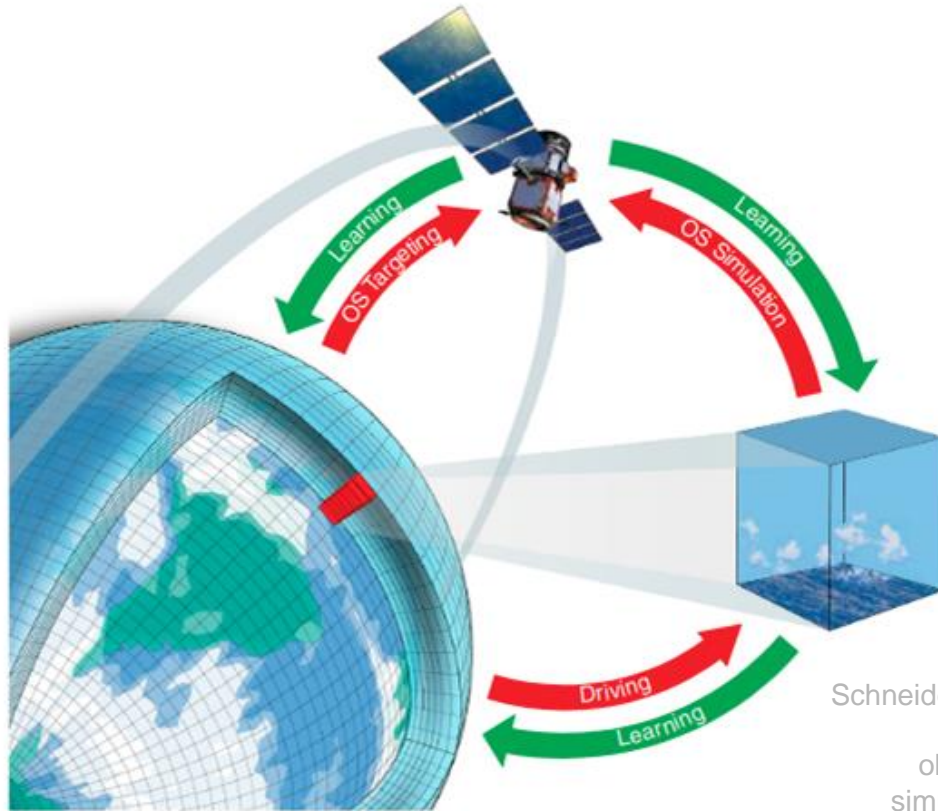
Publications

At conferences

In journals

Removing barriers

Building climate models that are ready to learn



Schneider, T., et al. "Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations." *Geophysical Research Letters* (2017)

Barriers to implementation

Machine learning

Climate science

What's exciting?

Big data!



Science!

Objectives

Well-defined is useful.

Broad is interesting.

Explainability

Second to prediction

Often the main goal

Data

Ideally clean and labelled

Many unlabeled features

Data formats

Images, csv, dataframes

Images, netcdf, misc

Data use

Integral to model



Data -> theory -> model

Existing code

Python, R, Julia



C/C++, Fortran

Publications

At conferences

In journals

Creating benchmark datasets

NASA Jet Propulsion Laboratory
California Institute of Technology

JPL HOME | EARTH | SOLAR SYSTEM | STARS & GALAXIES | SCIENCE & TECHNOLOGY

BRING THE UNIVERSE TO YOU: [Social Media Icons]

Airborne Visible / Infrared Imaging Spectrometer
AVIRIS
NEXT GENERATION

BENCHMARK METHANE DATA

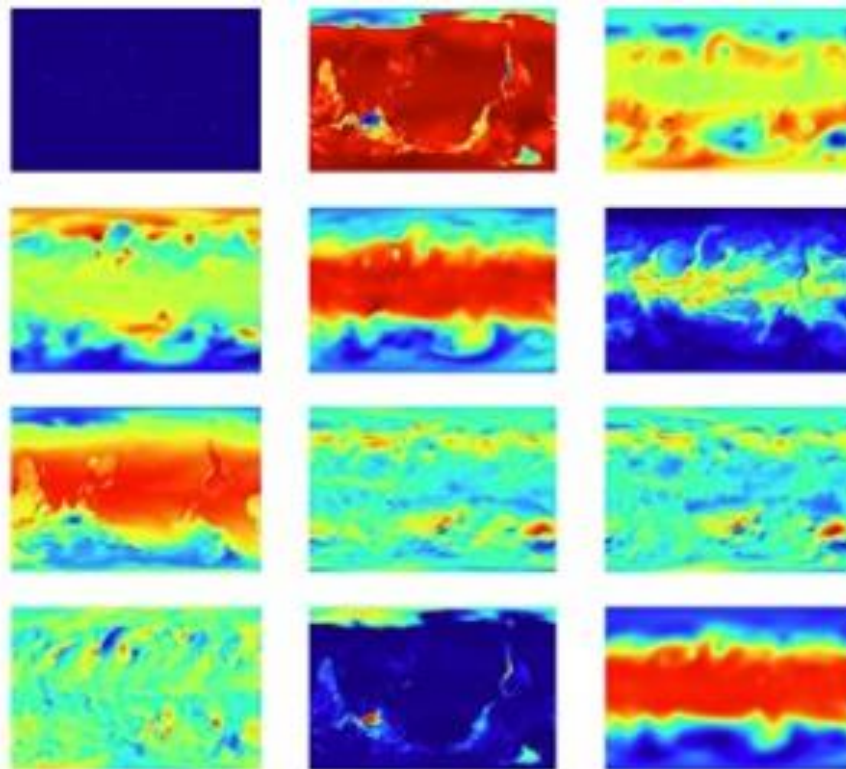
Home
Flight Status
AVIRIS-NG
AVIRIS Classic
Data
Benchmark Methane Data
Quicklooks
Publications
Links
Contact
News and Updates

Methane (CH₄) is a powerful greenhouse gas. CH₄ enters the atmosphere from various sources: naturally, from phenomena like geologic seeps; and artificially, through a range of processes like animal husbandry, decomposition in landfills, and oil and gas extraction and production. Institutions like the Jet Propulsion Laboratory are developing automated methods for the reliable detection, and potentially classification, of CH₄ sources from imaging spectrometer data. The key challenge is to distinguish CH₄ sources from background noise. Domain experts currently perform this task manually by visual inspection of the imaging spectrometer data. In consort with the NSF-sponsored Research Coordination Network on Intelligent Systems Research to Support Geosciences (IS-GEO, <https://is-geo.org/>), we have prepared a benchmark dataset to promote future research on automated detection. The dataset, designated JPL-CH₄-detection-2017-V1.0, contains sample imaging spectrometer data acquired in the "Four Corners" area - a hotspot of CH₄. This document describes the dataset to researchers developing and testing algorithms for CH₄ detection from imaging spectrometer data. The benchmark includes: (1) this description and (2) the dataset. The data was

The benchmark dataset is available online for use by researchers.

Access the AVIRIS-NG Methane Benchmark Data Set »

Download the Benchmark for CH₄ Source Detection Document



Barriers to implementation

Machine learning

Climate science

What's exciting?

Big data!



Science!

Objectives

Well-defined is useful.



Broad is interesting.

Explainability

Second to prediction



Often the main goal

Data

Ideally clean and labelled



Many unlabeled features

Data formats

Images, csv, dataframes



Images, netcdf, misc

Data use

Integral to model

Data -> theory -> model

Existing code

Python, R, Julia

C/C++, Fortran

Publications

At conferences

In journals

Running machine-learning oriented workshops



NeurIPS 2019 Workshop
December 13/14 in Vancouver, Canada

TACKLING CLIMATE CHANGE WITH MACHINE LEARNING

Submission deadline: September 11
Details at www.climatechange.ai

Organizers:
David Rolnick, Alexandre Lacoste, Tegan Maharaj, Priya Donti,
Lynn Kaack, John Platt, Jennifer Chayes, Yoshua Bengio

Barriers to implementation

Machine learning

Climate science

What's exciting?

Big data!



Science!

Objectives

Well-defined is useful.



Broad is interesting.

Explainability

Second to prediction



Often the main goal

Data

Ideally clean and labelled



Many unlabeled features

Data formats

Images, csv, dataframes

Images, netcdf, misc

Data use

Integral to model

Data -> theory -> model

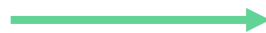
Existing code

Python, R, Julia

C/C++, Fortran

Publications

At conferences



In journals

Next steps

Learn more about machine learning

Online courses

- coursera.org/learn/machine-learning

Informational blogs

- towardsdatascience.com

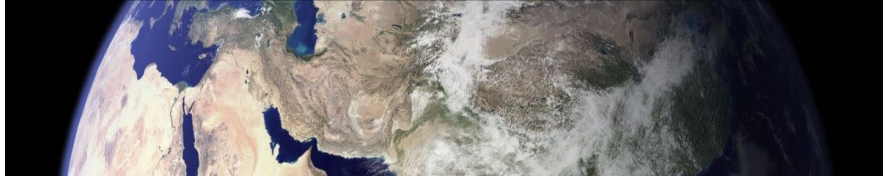
Python tutorials

- Scikit-learn: bit.ly/sklstrata, fastai: course.fast.ai

Learn more about machine learning for Earth, weather, and climate science

- McGovern, Amy, et al. *Bulletin of the American Meteorological Society* 98.10 (2017): 2073-2090. [Using artificial intelligence to improve real-time decision-making for high-impact **weather**.](#)
- Reichstein, Markus, et al. *Nature* 566.7743 (2019): 195. [Deep learning and process understanding for data-driven **Earth system science**.](#)
- Karpatne, Anuj, et al. *IEEE Transactions on Knowledge and Data Engineering* (2018). [Machine learning for the **geosciences**: Challenges and opportunities.](#)
- Gil, Y., Pierce, S. A., ... & Horel, J. (2018). *Communications of the ACM*, 62(1), 76-84. [Intelligent systems for **geosciences**: an essential research agenda.](#)
- Rolnick, D., Donti, P., Kaack, L., Kochanski, K., et al. *arXiv preprint arXiv:1906.05433* (2019). [Tackling **climate change** with machine learning](#)

Make connections



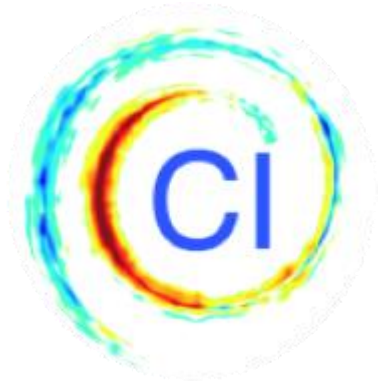
Climate Change AI
climatechange.ai

Make connections



Climate Change AI

climatechange.ai



Climate
Informatics
climateinformatics.org

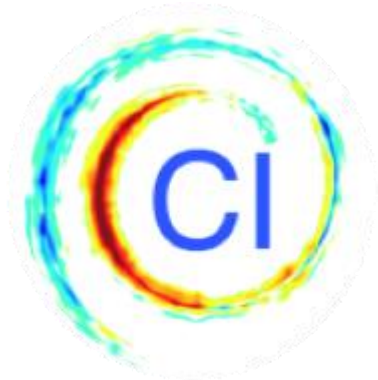
Make connections



Climate Change AI
climatechange.ai

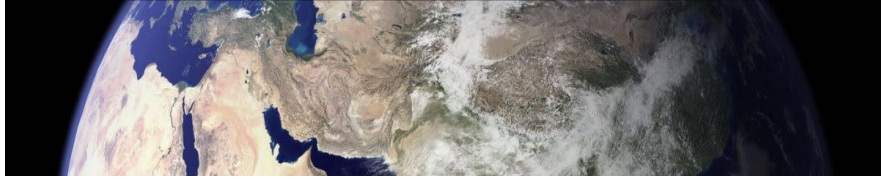


Intelligent Systems and
Geosciences
is-geo.org



Climate
Informatics
climateinformatics.org

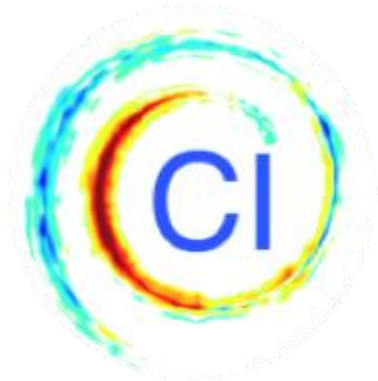
Make connections



Climate Change AI
climatechange.ai



Intelligent Systems and
Geosciences
is-geo.org



Climate
Informatics
climateinformatics.org



AMS Committee on AI
for Env. Science

Thanks

Greg Tucker, David Rolnick, Ghaleb Abdulla, Divya Mohan,
Jenna Horrall, Priya Donti, Surya Karthik Mukkavilli,
Barry Rountree, Goodwin Gibbons



A satellite-style image of the Earth, showing the Middle East, North Africa, and parts of Europe and Asia. The word "Questions?" is overlaid in white text in the upper center.

Questions?