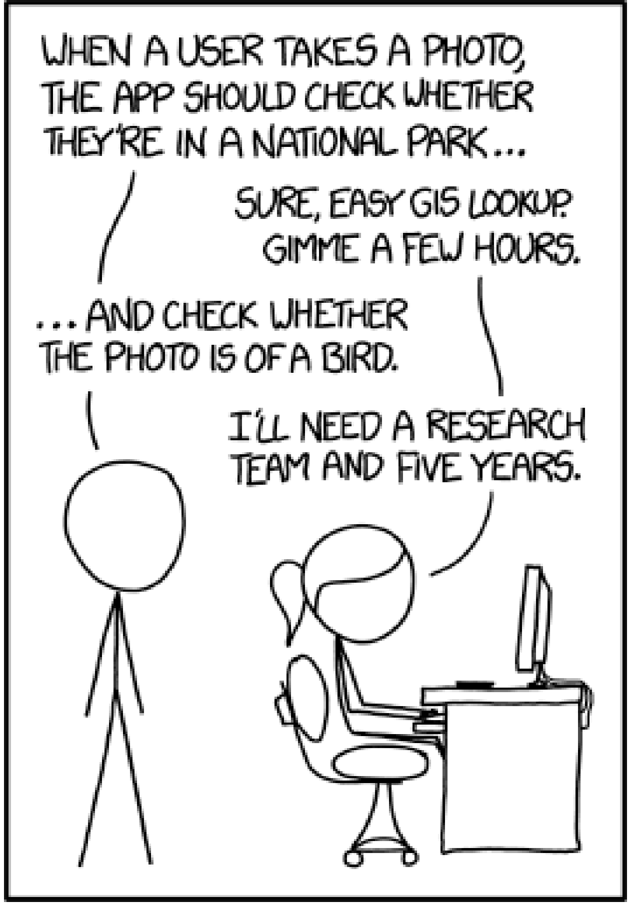
Accelerating Earth and climate modeling with machine learning

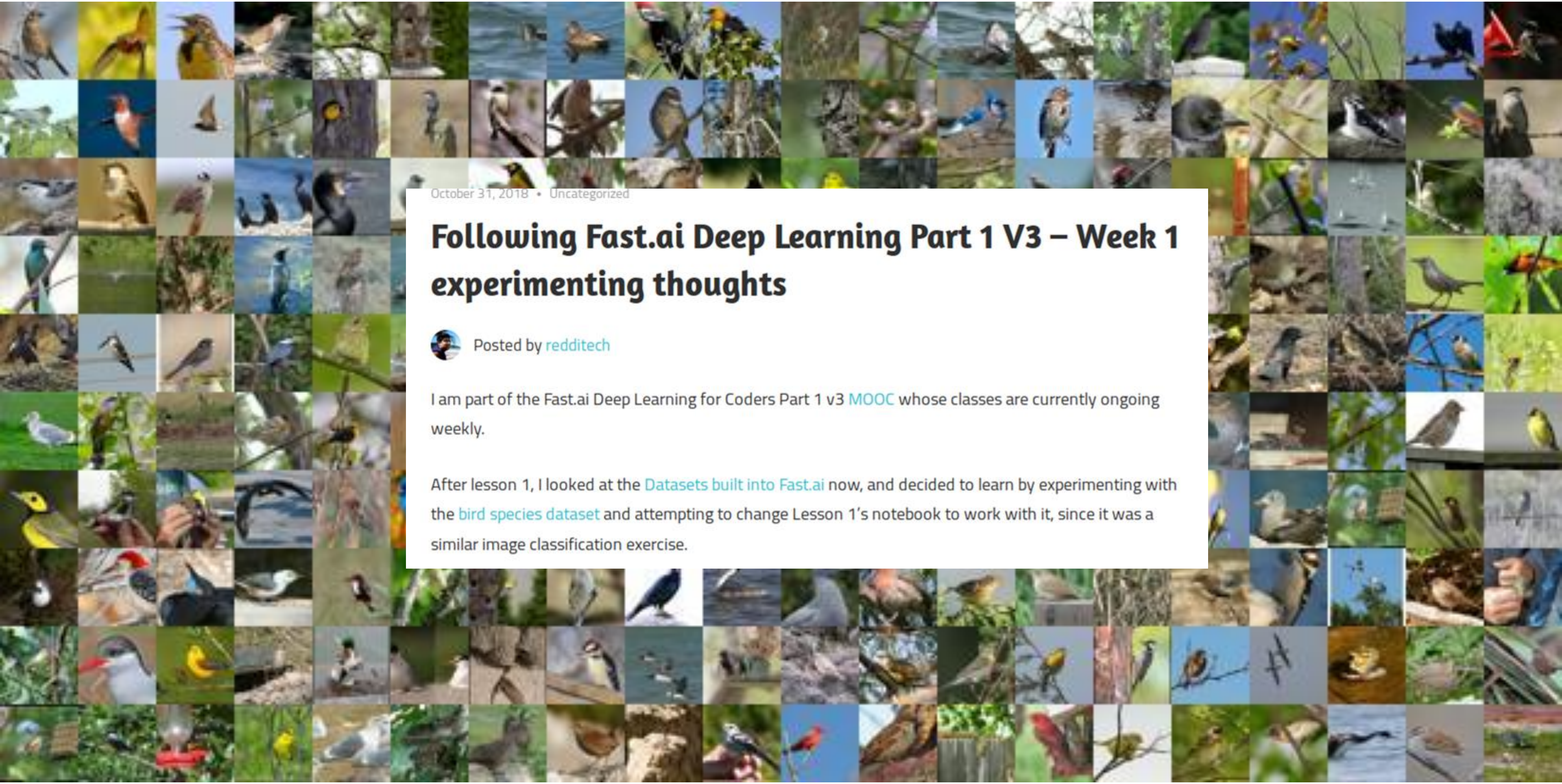
Kelly Kochanski
NCAR Multicore Workshop 2019



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

2014

xkcd.com/1425/



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?

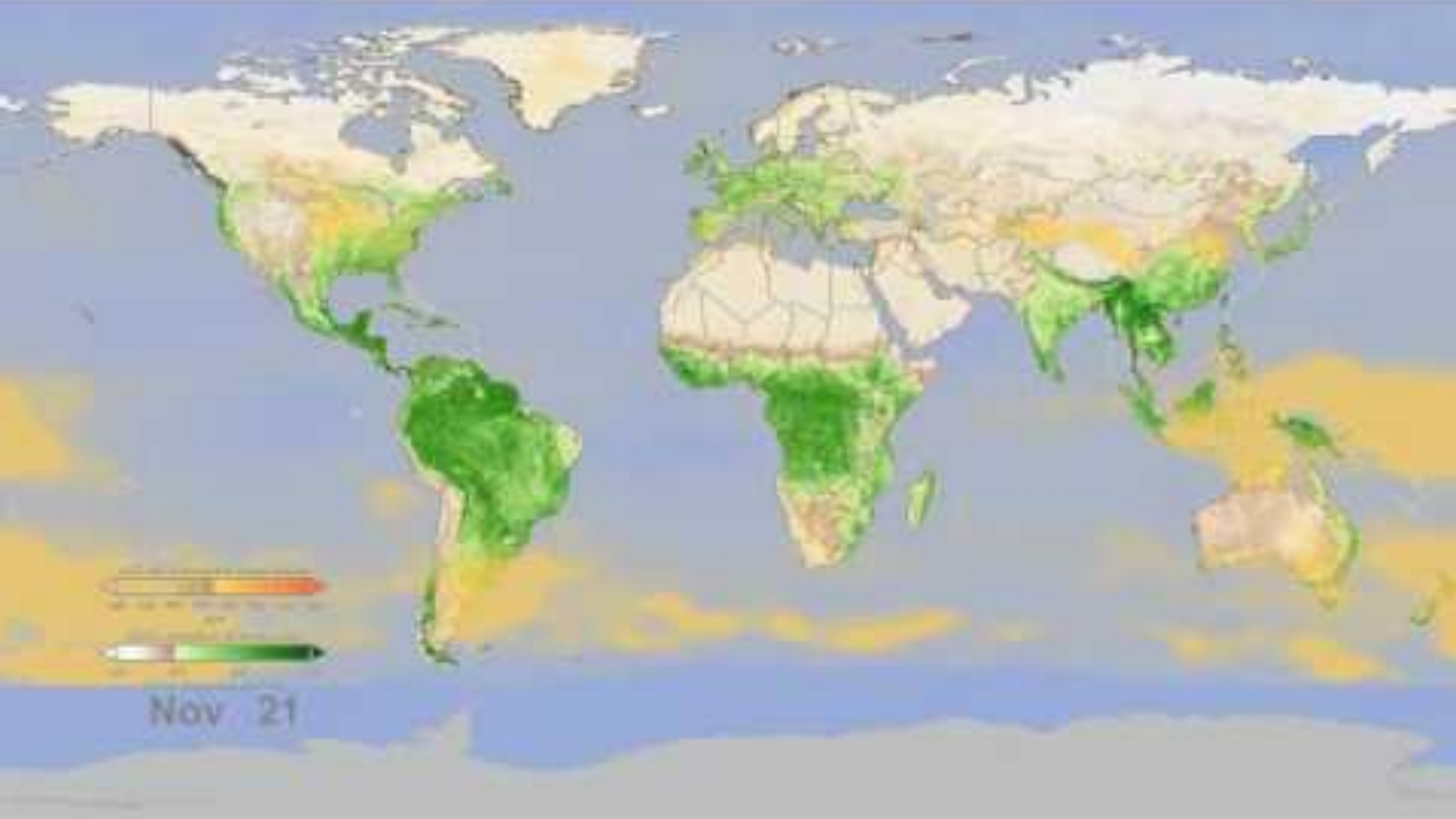
Machine learning at its most basic is the practice of using algorithms to parse data, learn from it, and then make a determination or prediction about something in the world.

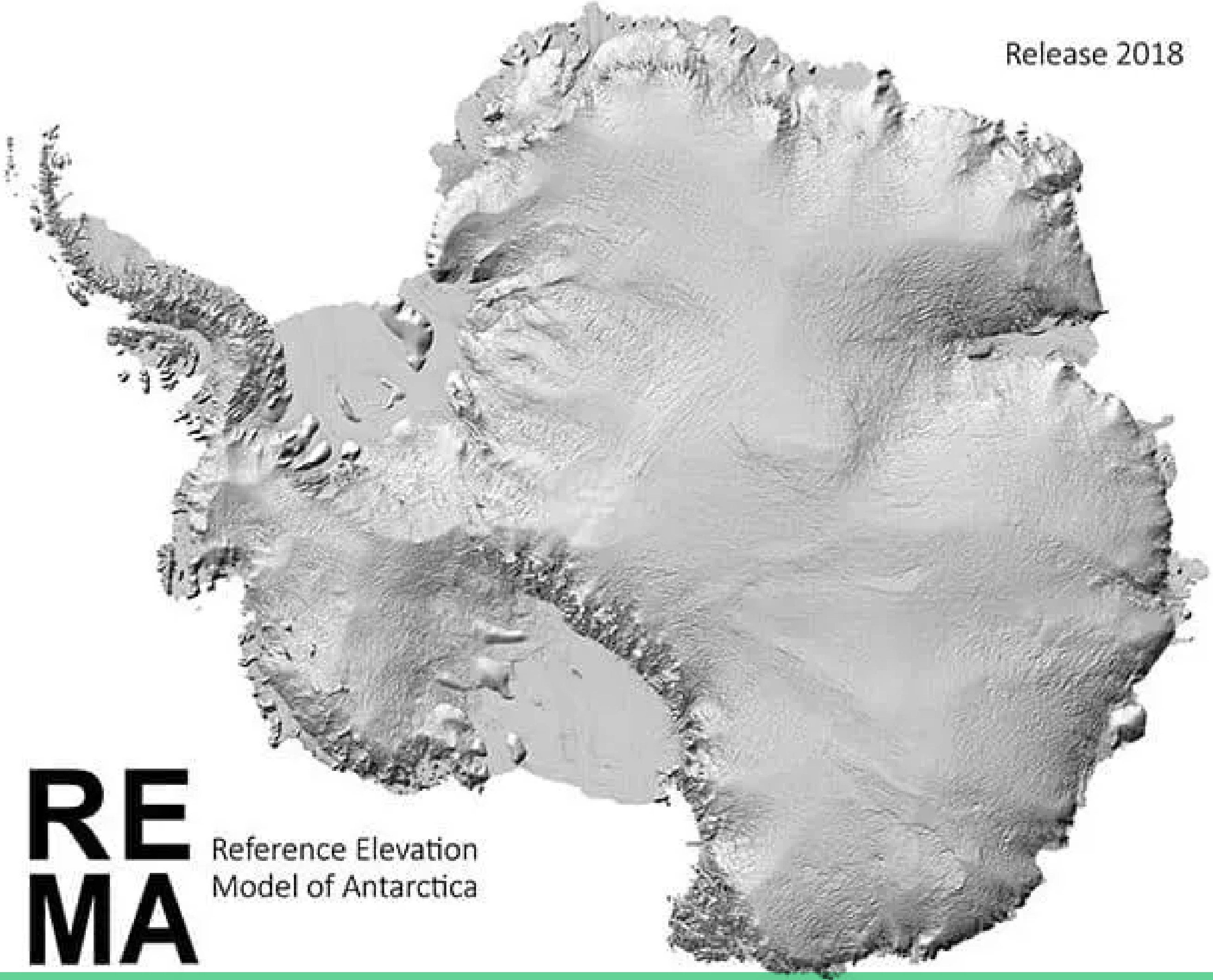
Michael Copeland
2016

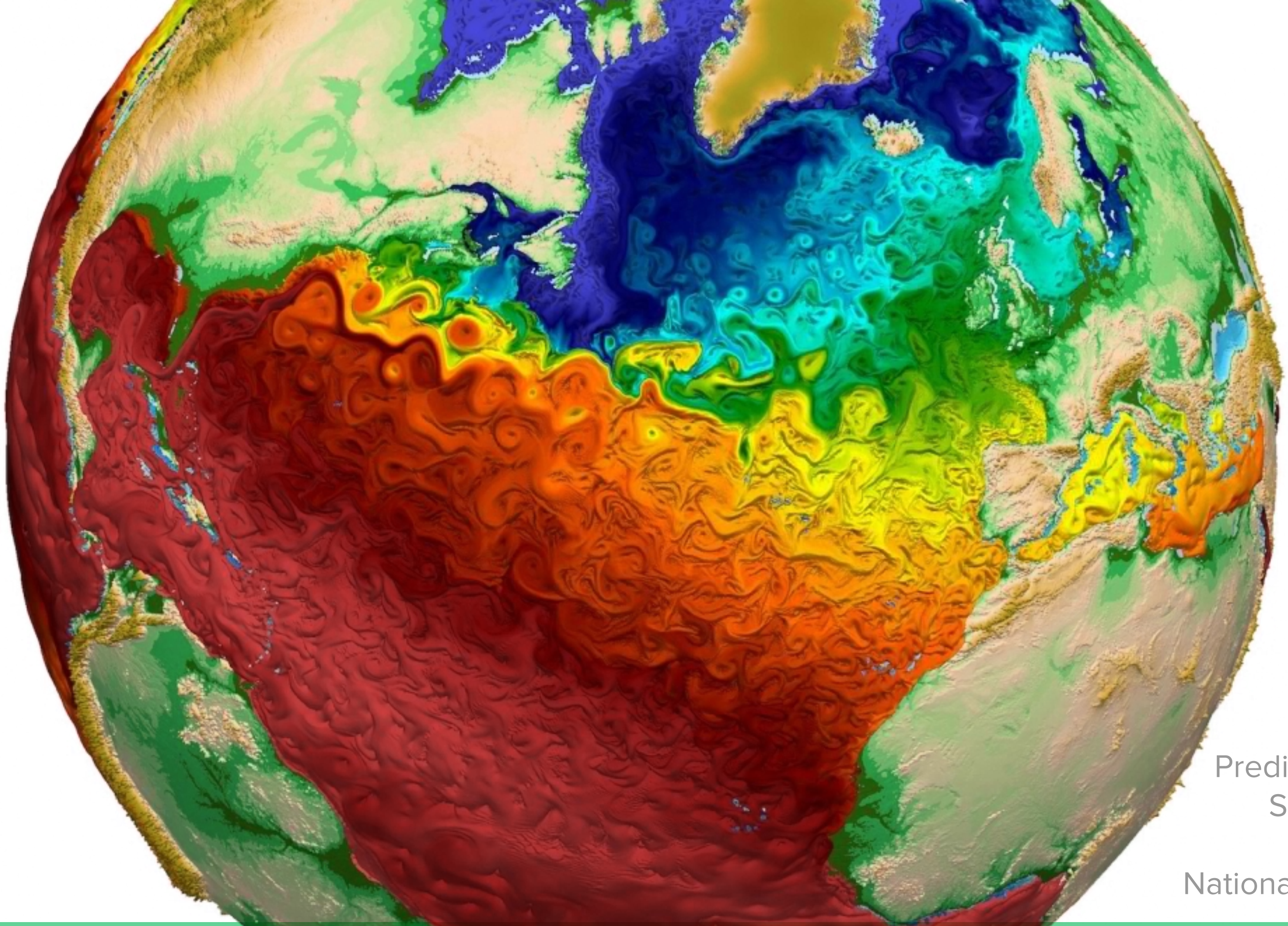
Why is machine learning relevant to Earth System Modeling now?

Current trends 1/3

Machine learning offers solutions to once-intractable problems


Nov 21


Release 2018
REMA
Reference Elevation
Model of Antarctica

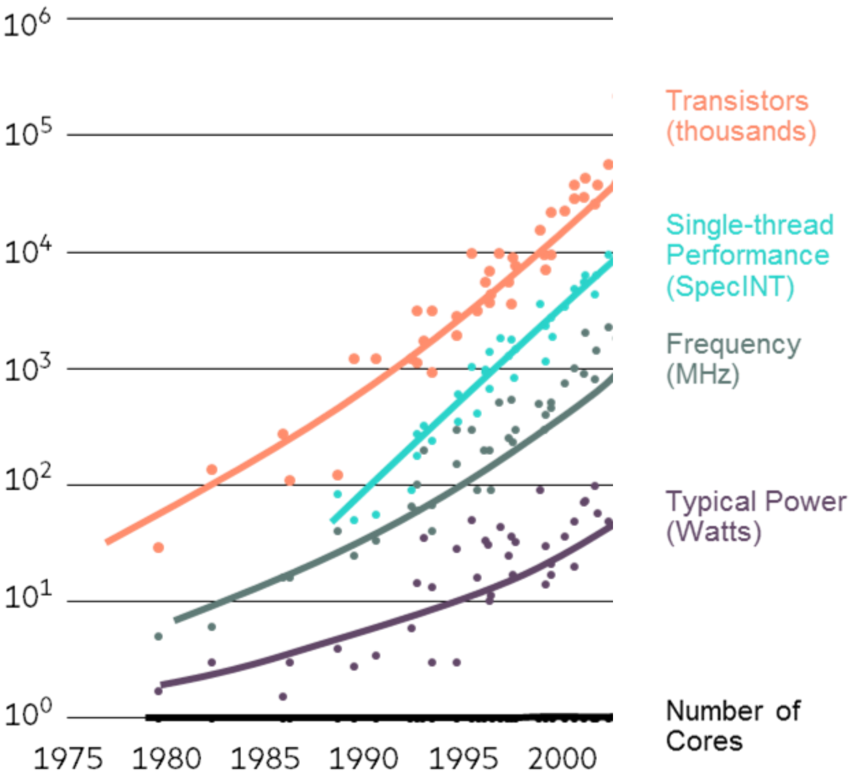


Model for
Prediction Across
Scales (2015),
Los Alamos
National Laboratory

Current trends 2/3

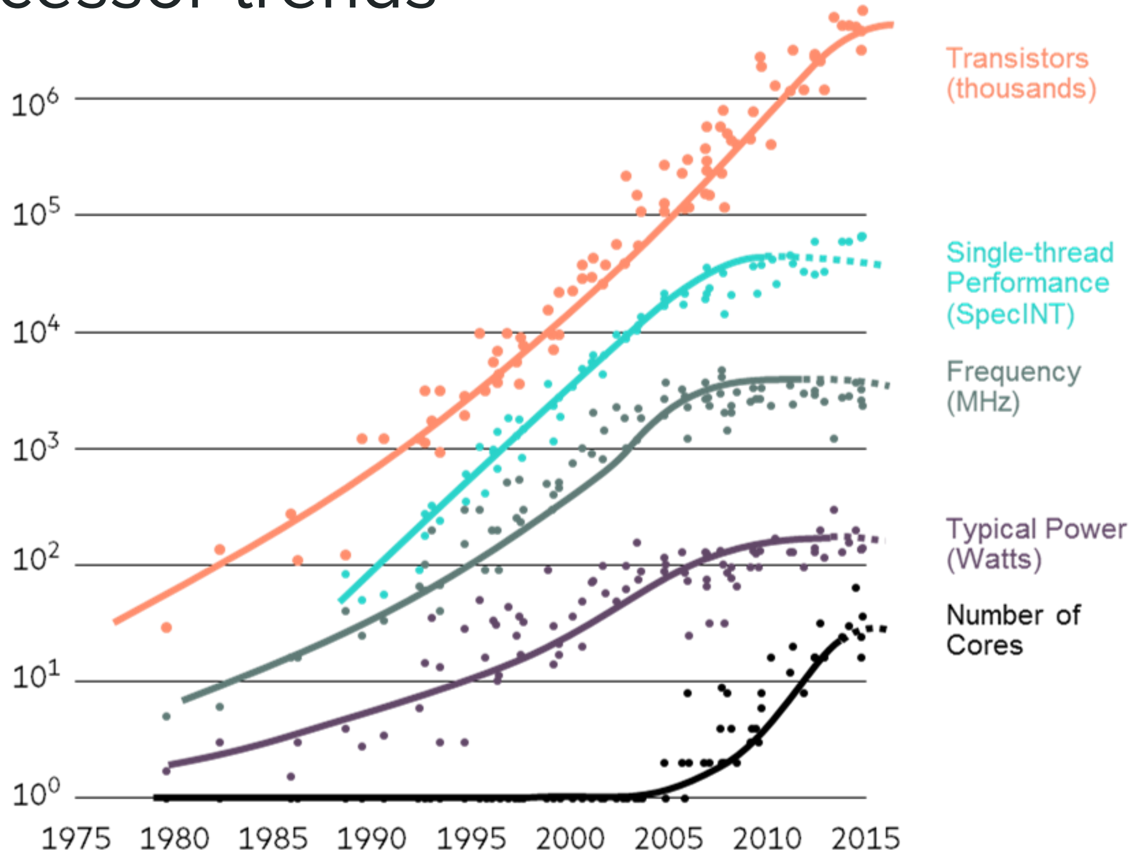
New data streams increase the potential power of data-driven models

Microprocessor trends



Karl Rupp, World Economic Forum, 2018

Microprocessor trends



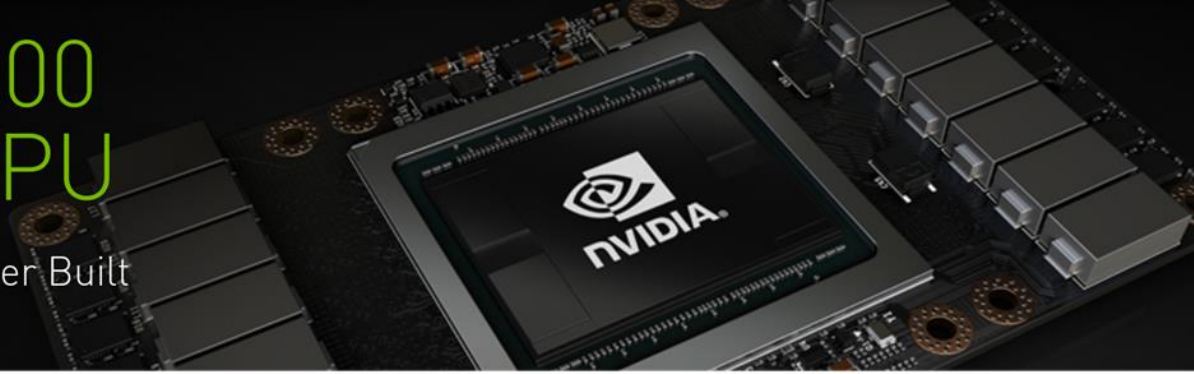
Karl Rupp, World Economic Forum, 2018

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. Donti[2], Lynn H. Kaack[3], Kelly Kochanski[4], Alexandre Lacoste[5], Kris Sankaran[6,7], Andrew Slavin Ross[8], Nikola Milojevic-Dupont[9,10], Natasha Jaques[11], Anna Waldman-Brown[11], Alexandra Luccioni[6,7], Tegan Maharaj[6,7], Evan D. Sherwin[2], S. Karthik Mukkavilli[6,7], Konrad P. Kording[1], Carla Gomes[12], Andrew Y. Ng[13], Demis Hassabis[14], John C. Platt[15], Felix Creutzig[9,10], Jennifer Chayes[16], Yoshua Bengio[6,7]

[1]University of Pennsylvania, [2]Carnegie Mellon University, [3]ETH Zürich, [4]University of Colorado Boulder, [5]Element AI, [6]Mila, [7]Université de Montréal, [8]Harvard University, [9]Mercator Research Institute on Global Commons and Climate Change, [10]Technische Universität Berlin, [11]Massachusetts Institute of Technology, [12]Cornell University, [13]Stanford University, [14]DeepMind, [15]Google AI, [16]Microsoft Research

climatechange.ai

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

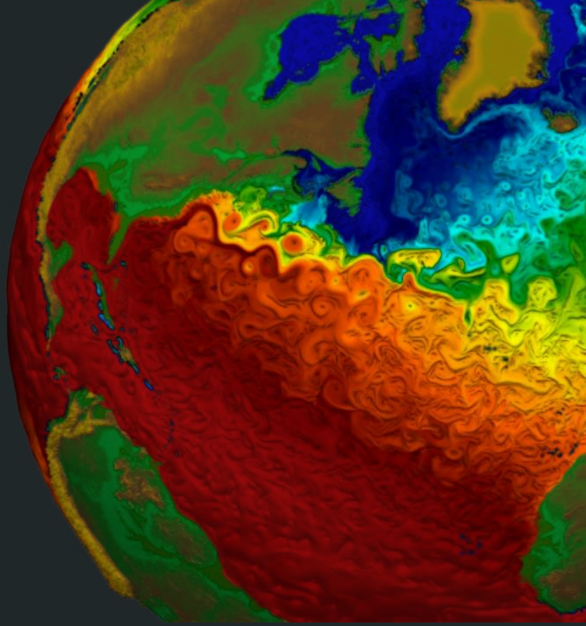
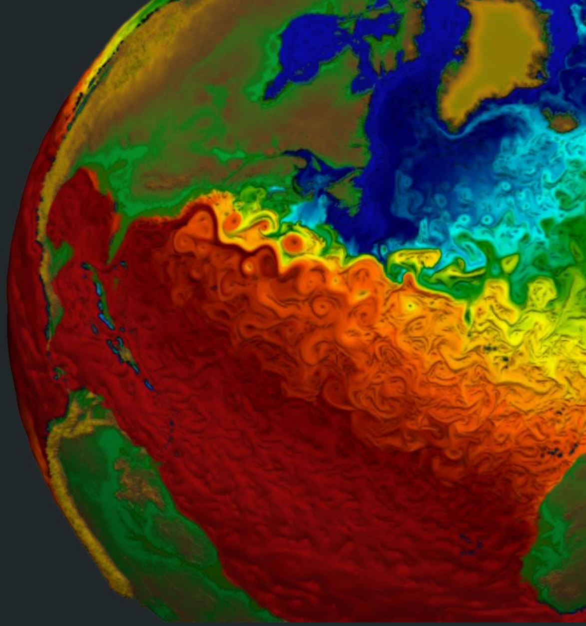


Image: MPAS-Ocean
Los Alamos National Lab

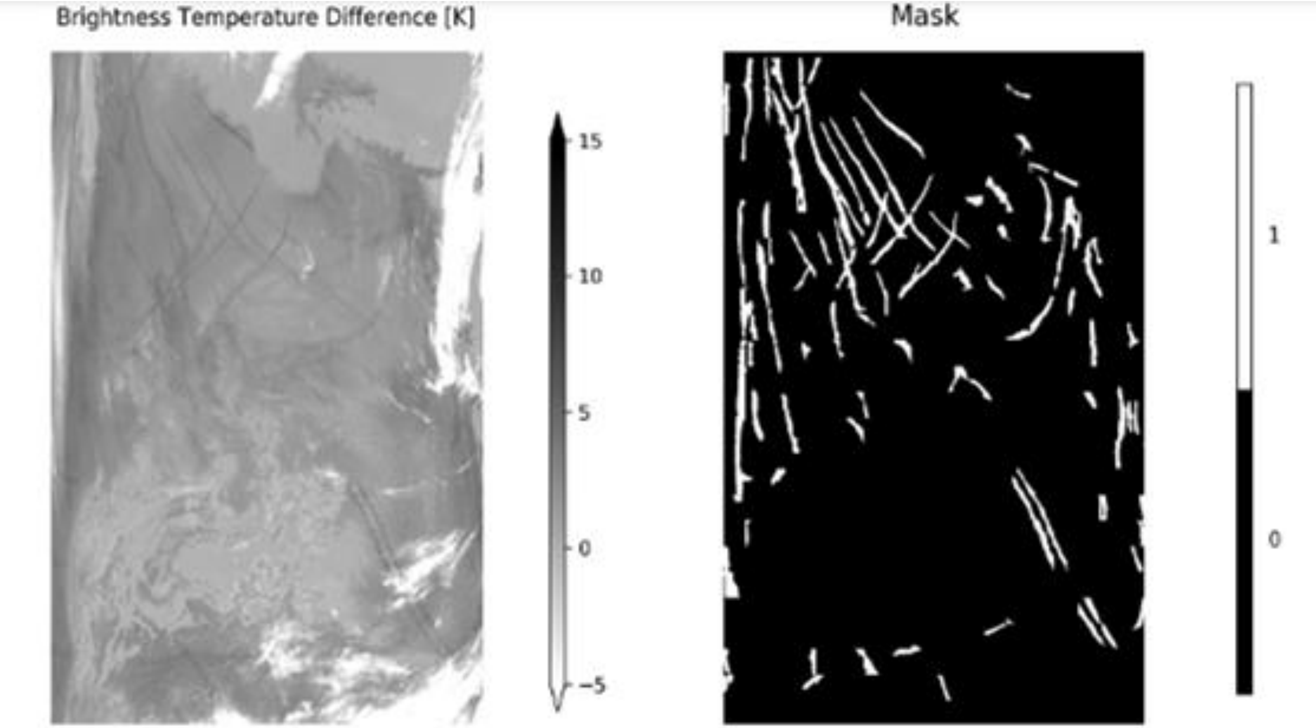
How can we use machine learning to build better Earth System Models?



Aims:

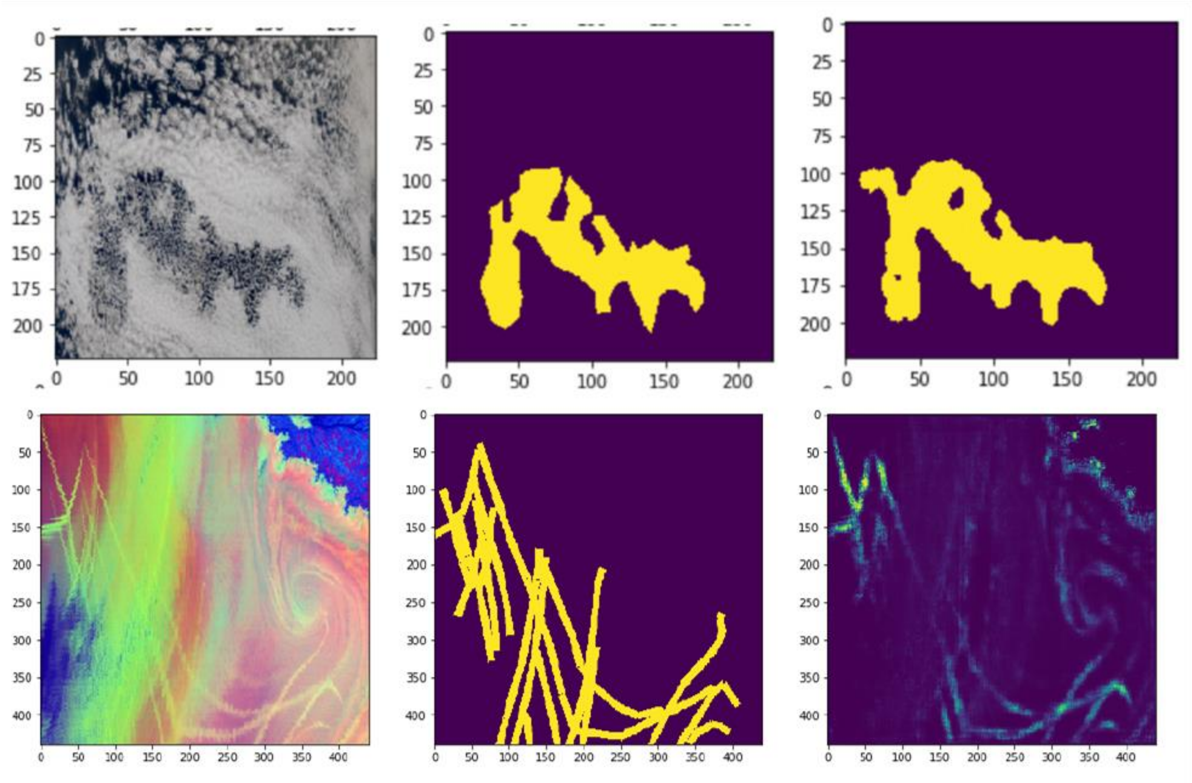
- To solve long-standing problems with new methods
- To integrate new sources of data into existing models
- To take advantage of new computing hardware

Monitoring marine clouds



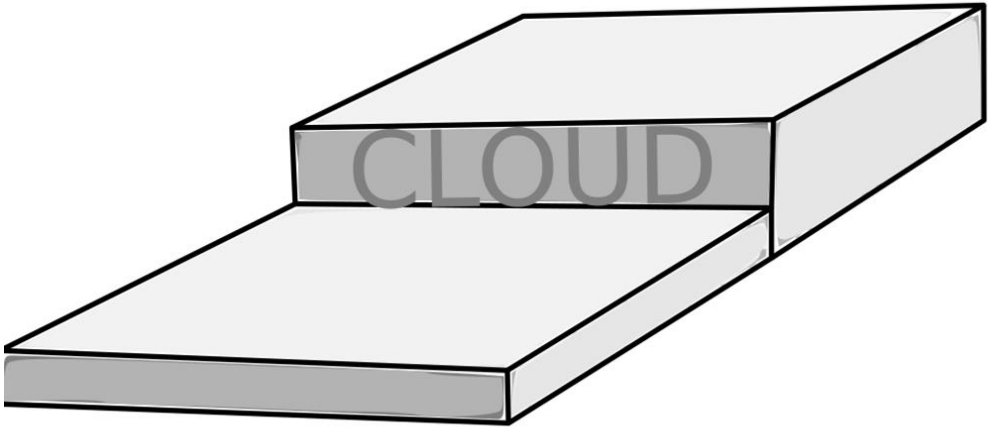
Yuan, Tianle, et al. "Automatically Finding Ship-tracks to Enable Large-scale Analysis of Aerosol-Cloud Interactions." *Geophysical Research Letters* (2019).

Monitoring marine clouds

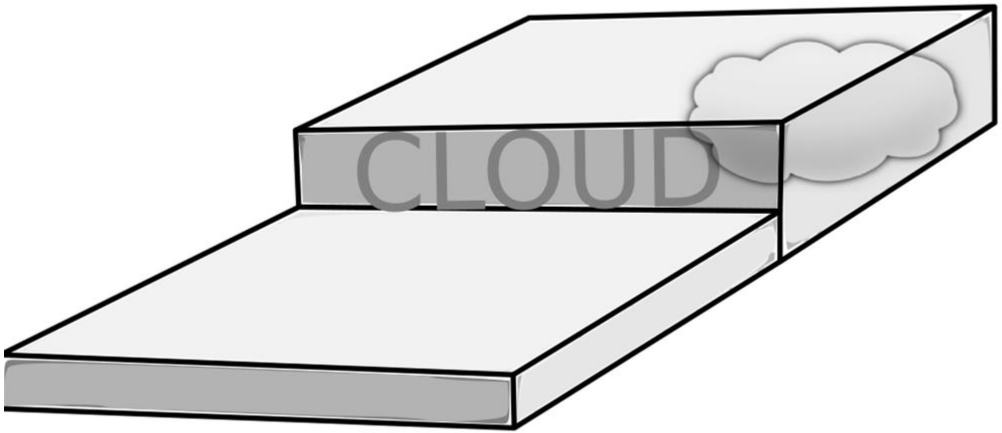


Watson-Parris, Duncan, et al. "Detecting anthropogenic cloud perturbations with deep learning." International Conference on Machine Learning, 2019.

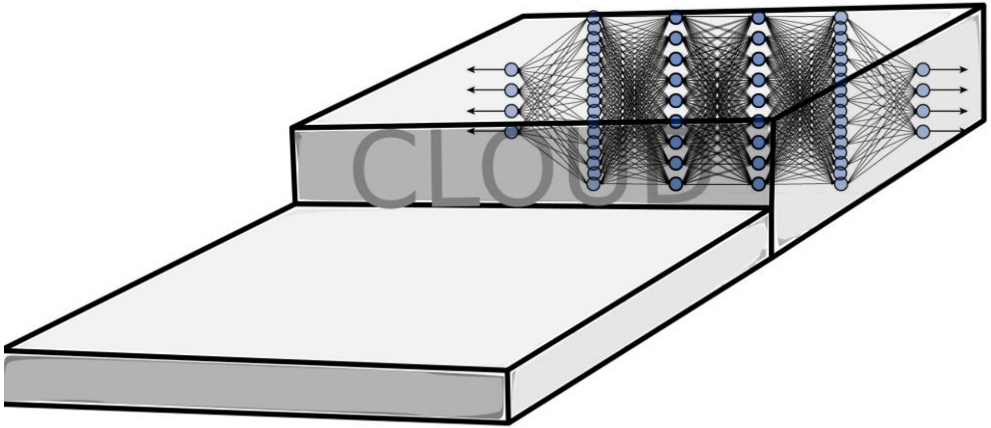
Improving convection + aerosol modelling



Improving convection + aerosol modelling

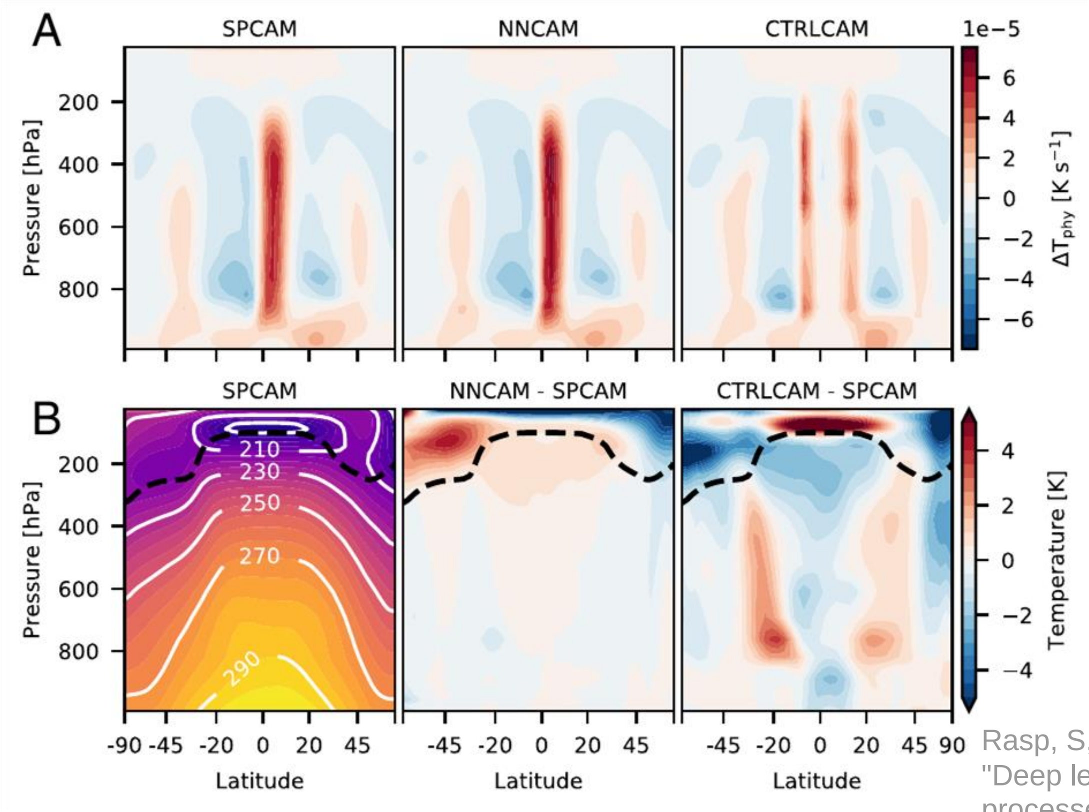


Improving convection + aerosol modelling



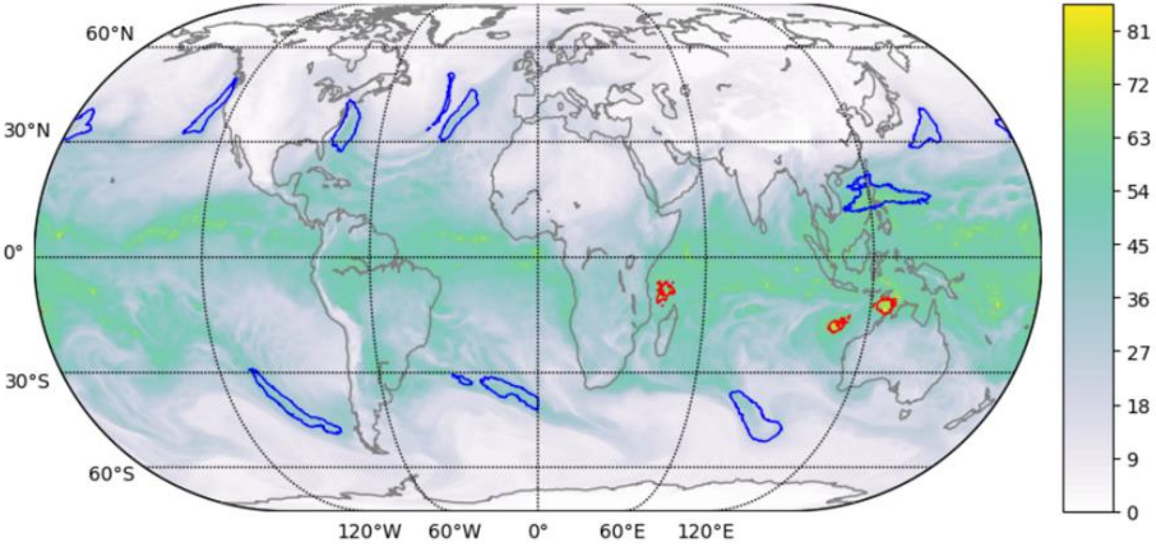
Gentine, Pierre, et al. "Could machine learning break the convection parameterization deadlock?." *Geophysical Research Letters* 45.11 (2018): 5742-5751.

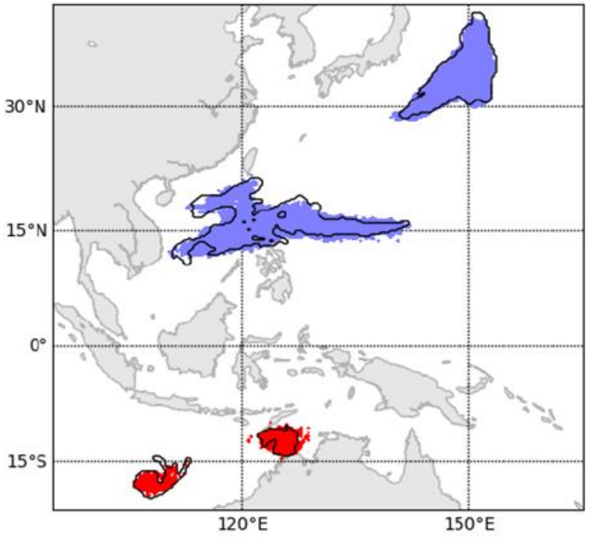
Improving convection + aerosol modelling



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

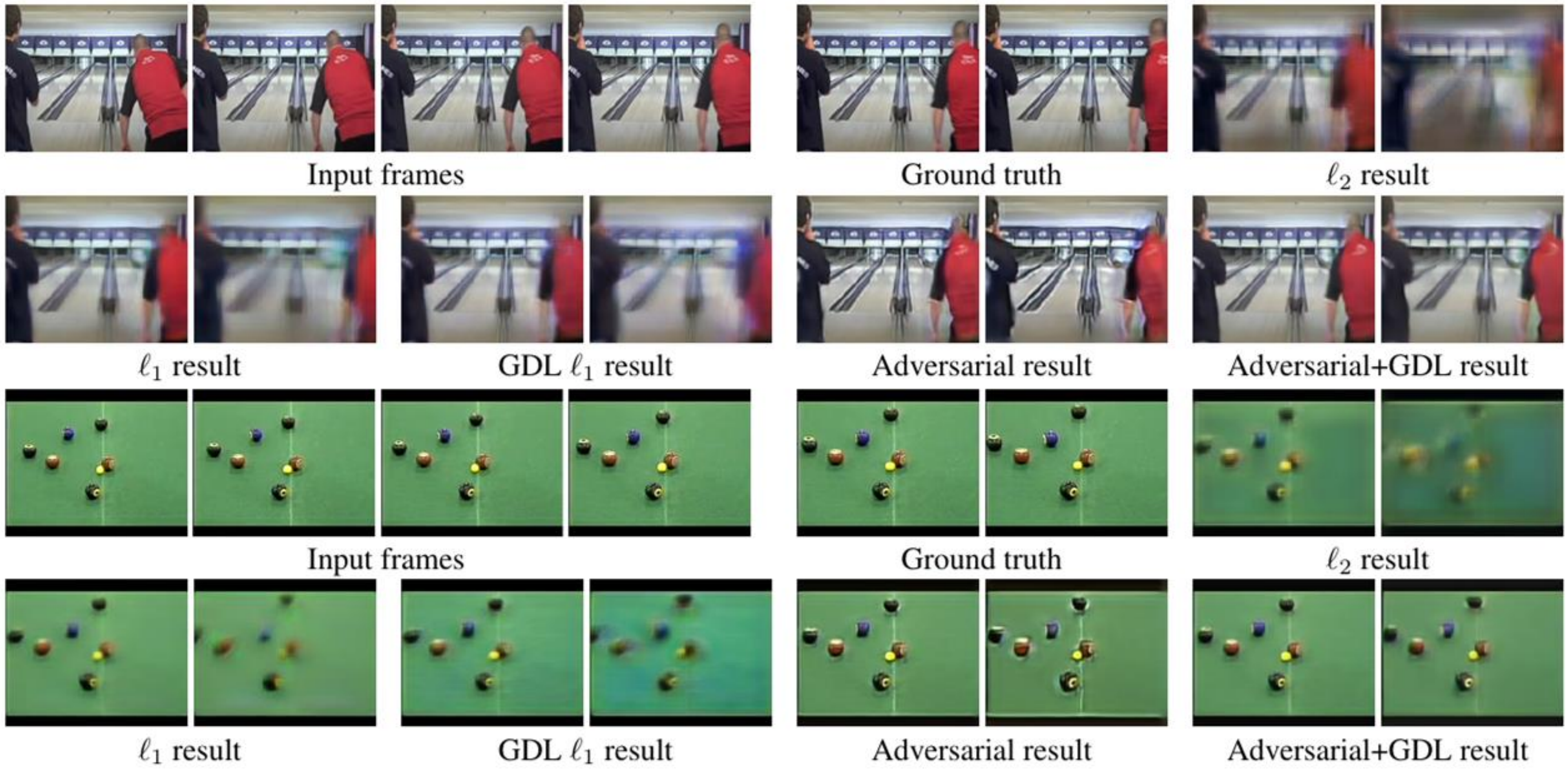
Tracking extreme events





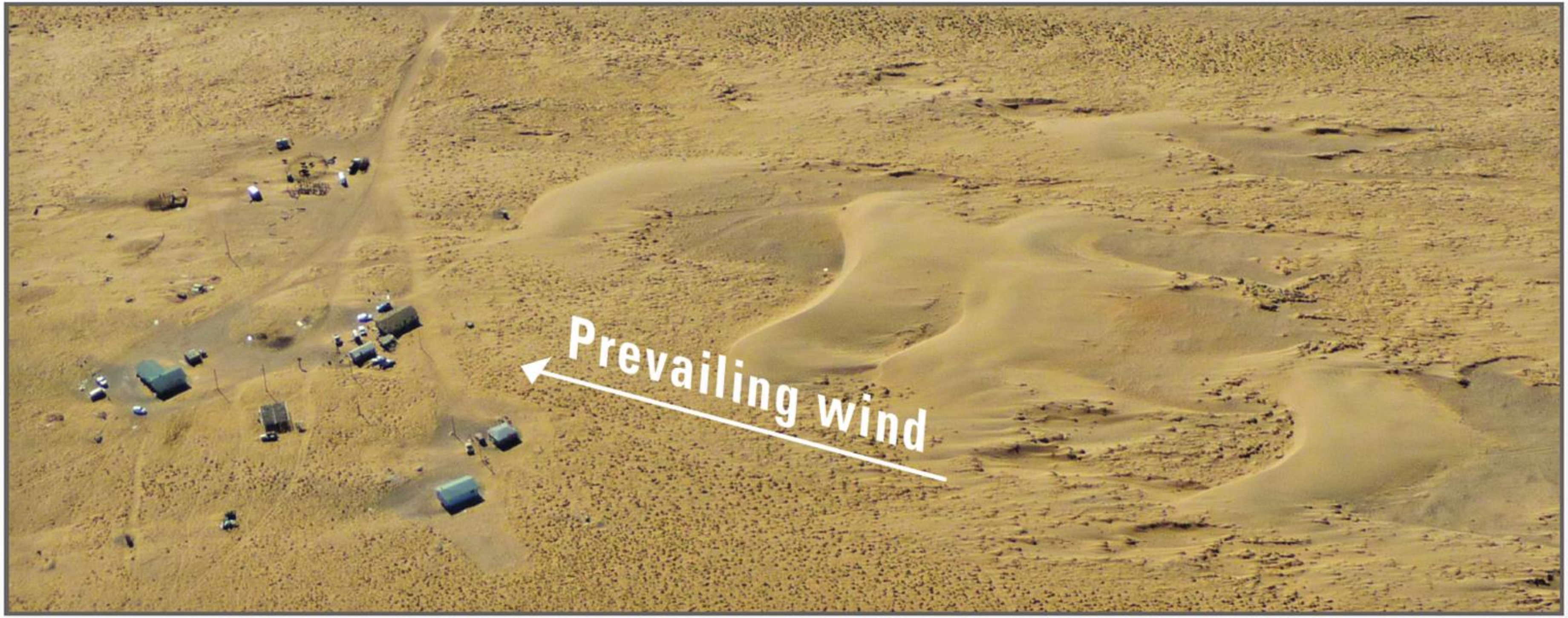
Kurth, Thorsten, et al. "Exascale deep learning for climate analytics." *Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis*. IEEE Press, 2018.

Deep learning for spatio-temporal patterns



Mathieu, Michael, Camille Couprie, and Yann LeCun. "Deep multi-scale video prediction beyond mean square error." (2016)

Deep learning for spatio-temporal patterns

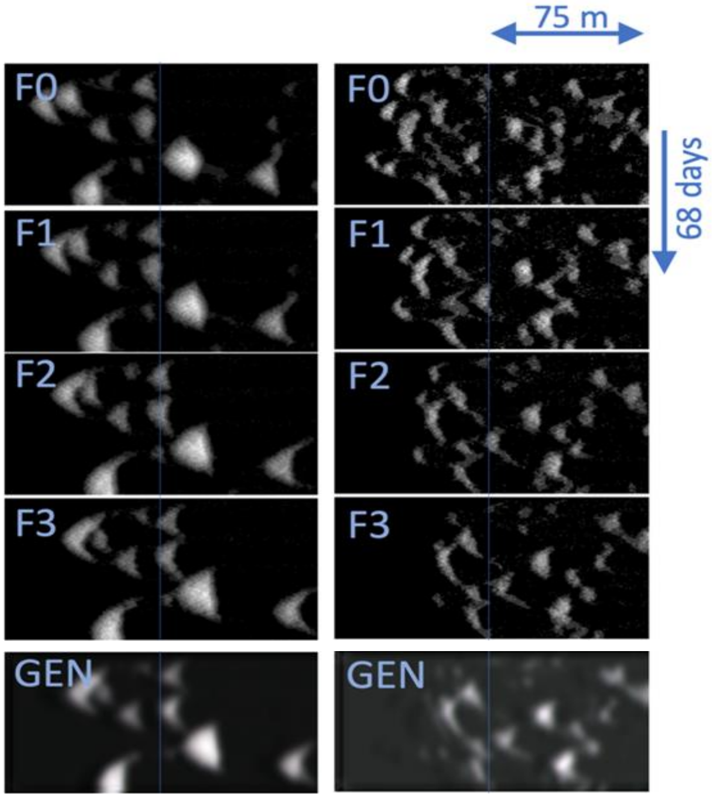


Reedster, Mogle and Bogel, 'Monitoring and analysis of sand dune movement and growth on the Navajo Nation, Southwestern United States' (2011) USGS Fact Sheet 3085.

Deep learning for spatio-temporal patterns

Simulated example

Generated frame(s)



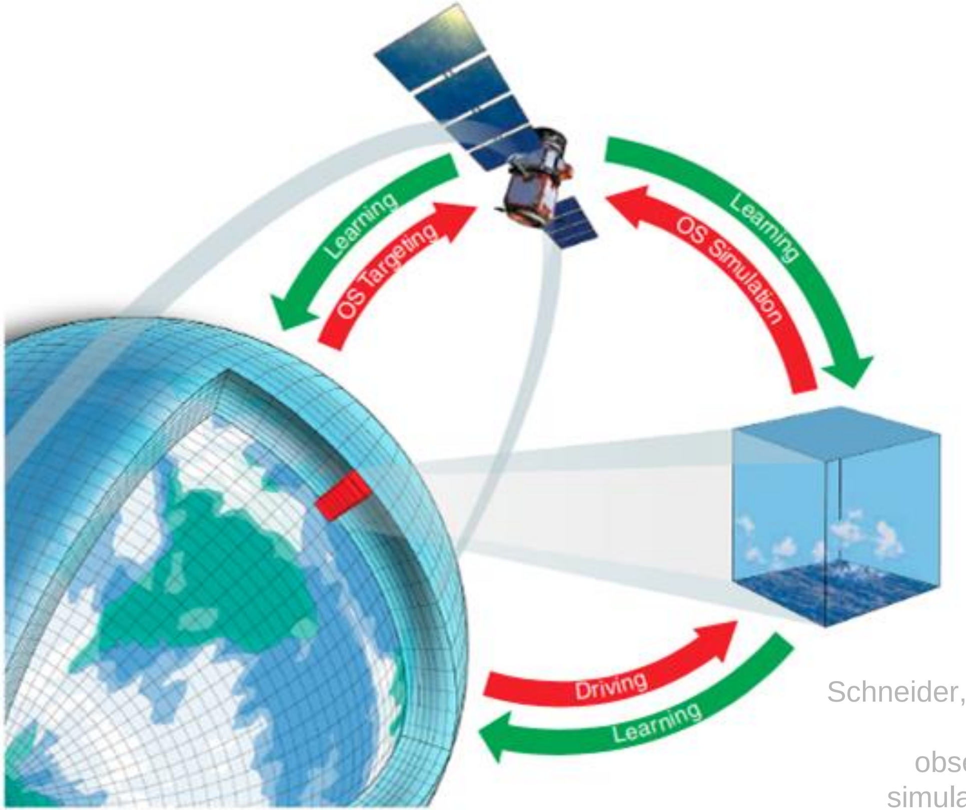
Barriers to implementation

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

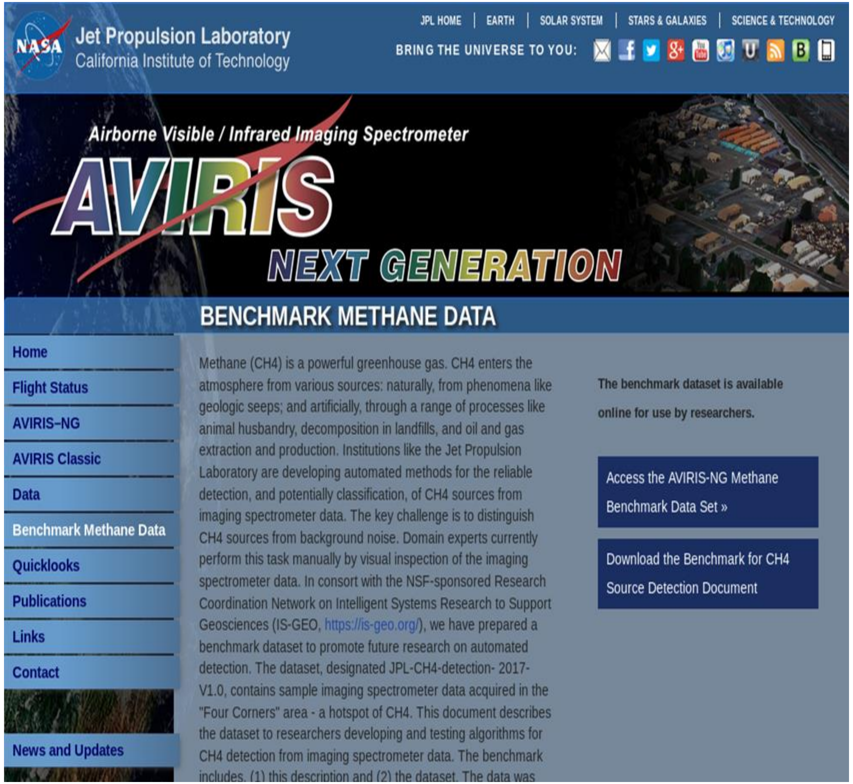


Schneider, T., et al. "Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations." *Geophysical Research Letters* (2017)

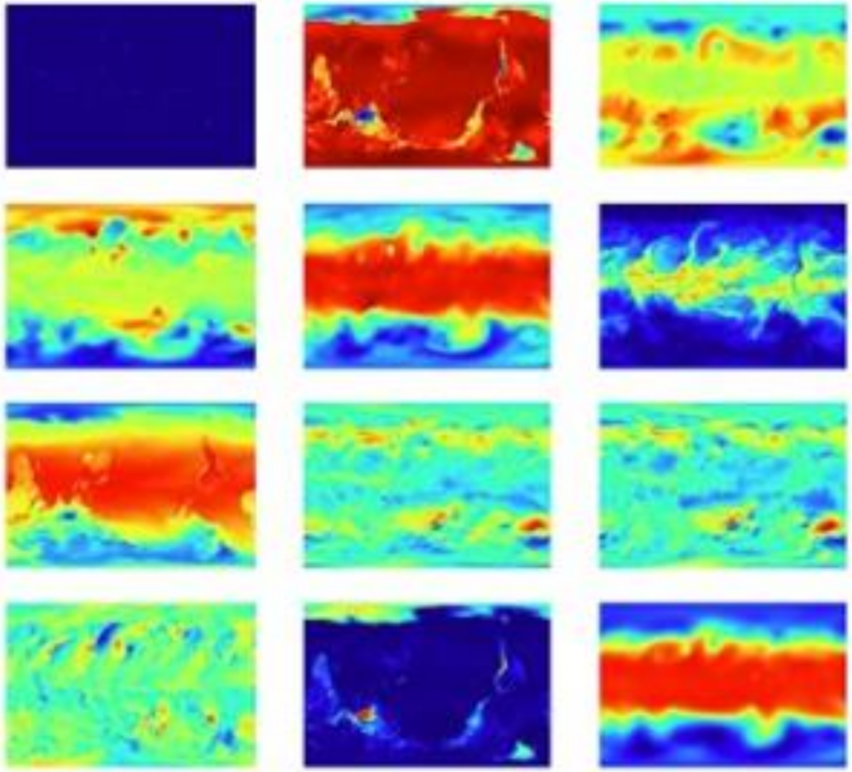
Barriers to implementation

	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



is-geo.org/benchmarks: JPL-CH4-detection-2017-V1.0



extremeweatherdataset.github.io

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

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

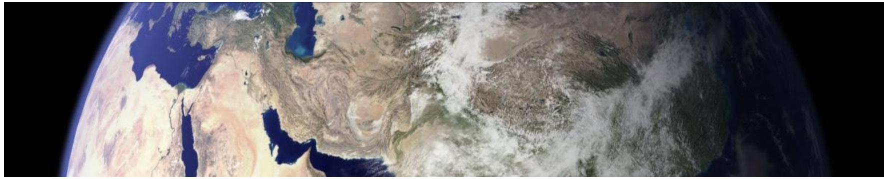
Python tutorials

- 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

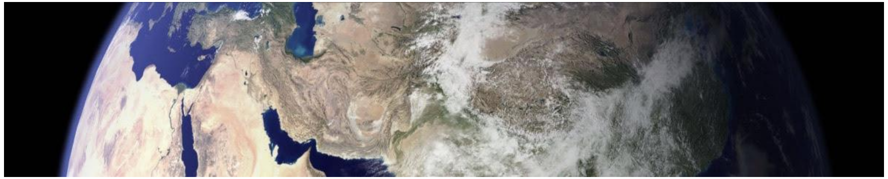
Make connections



Climate Change AI

climatechange.ai

Make connections



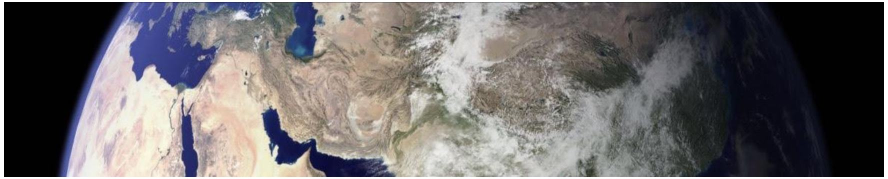
Climate Change AI

climatechange.ai

Climate Informatics

climateinformatics.org

Make connections

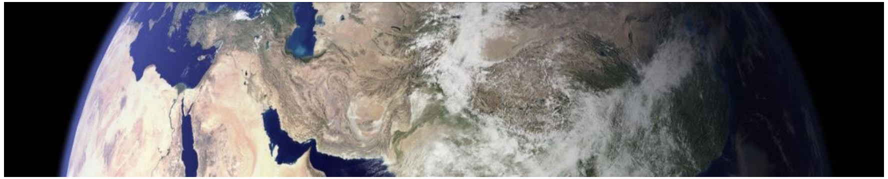


Climate Change AI
climatechange.ai

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

DOE CSGF


Questions?