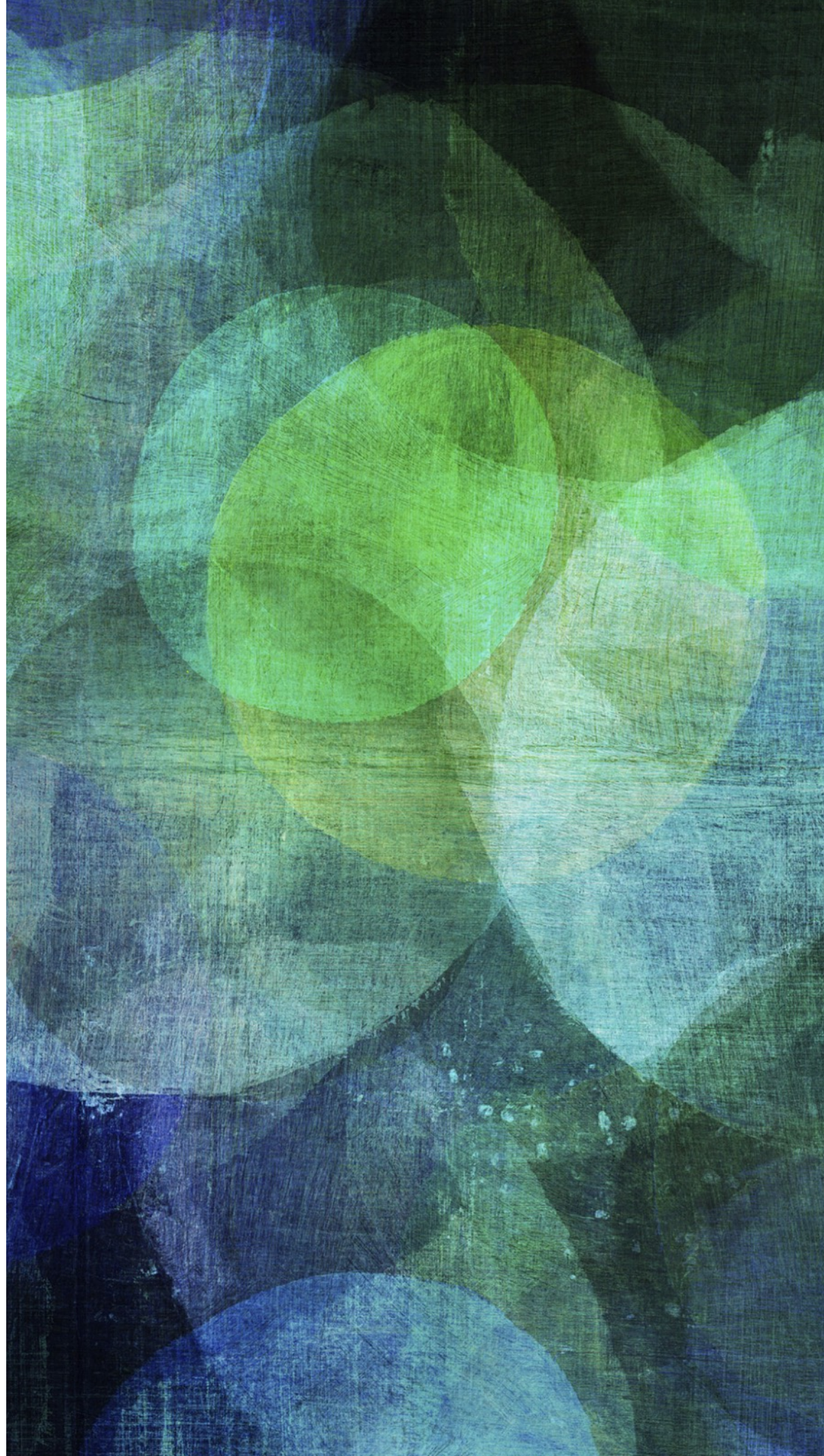
Machine learning with subsurface datasets—current state of the science and practice

Paul Johnson

Los Alamos NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY Office of Fossil Energy

U.S. DEPARTMENT OF ENERGY | Office of Science



CONTENTS

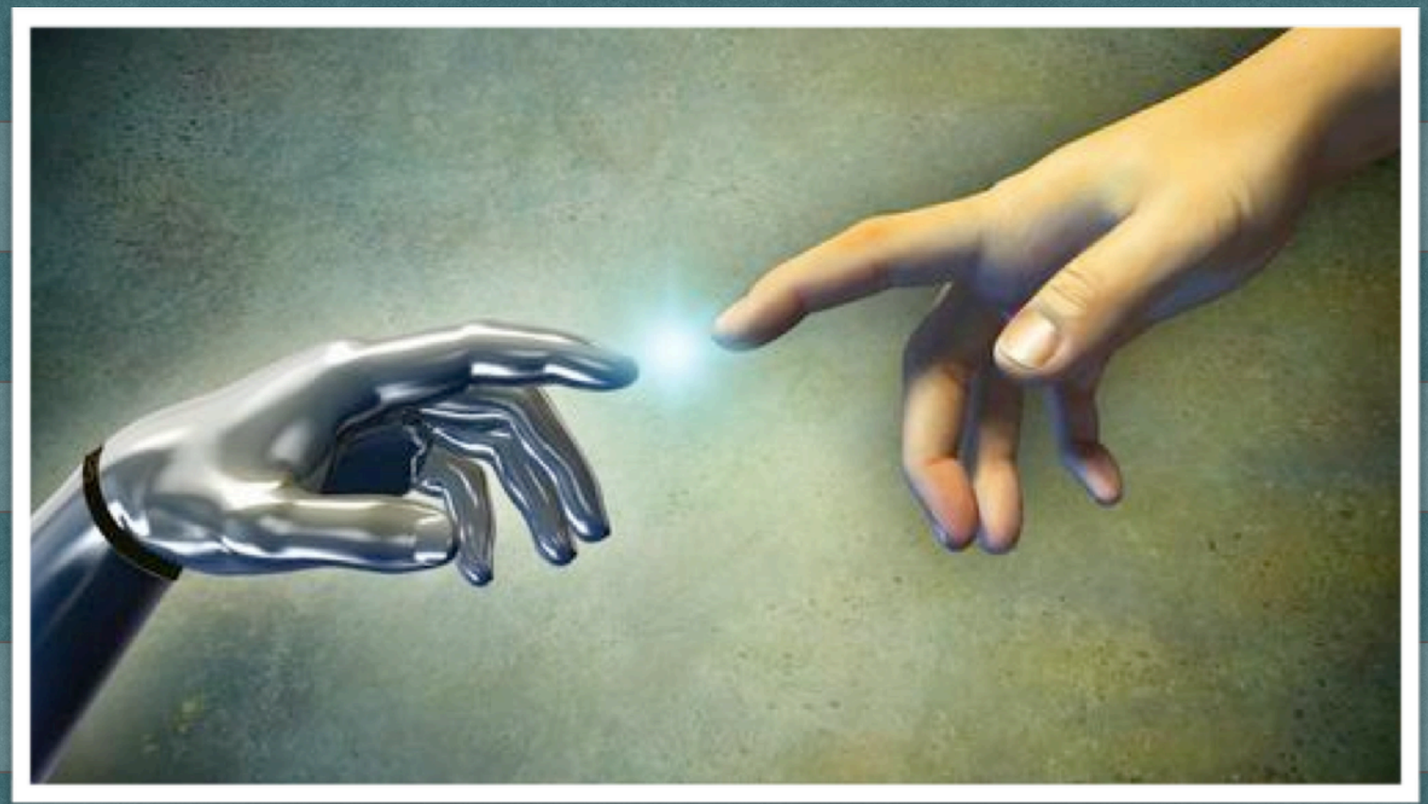
- Brief introduction and background to machine learning
- Supervised learning applied to laboratory fault physics
- Cascadia
- Perspective: path to future

TECHNOLOGICAL SHIFTS LEADING TO REVOLUTIONARY ADVANCES IN SOLID EARTH GEOSCIENCE

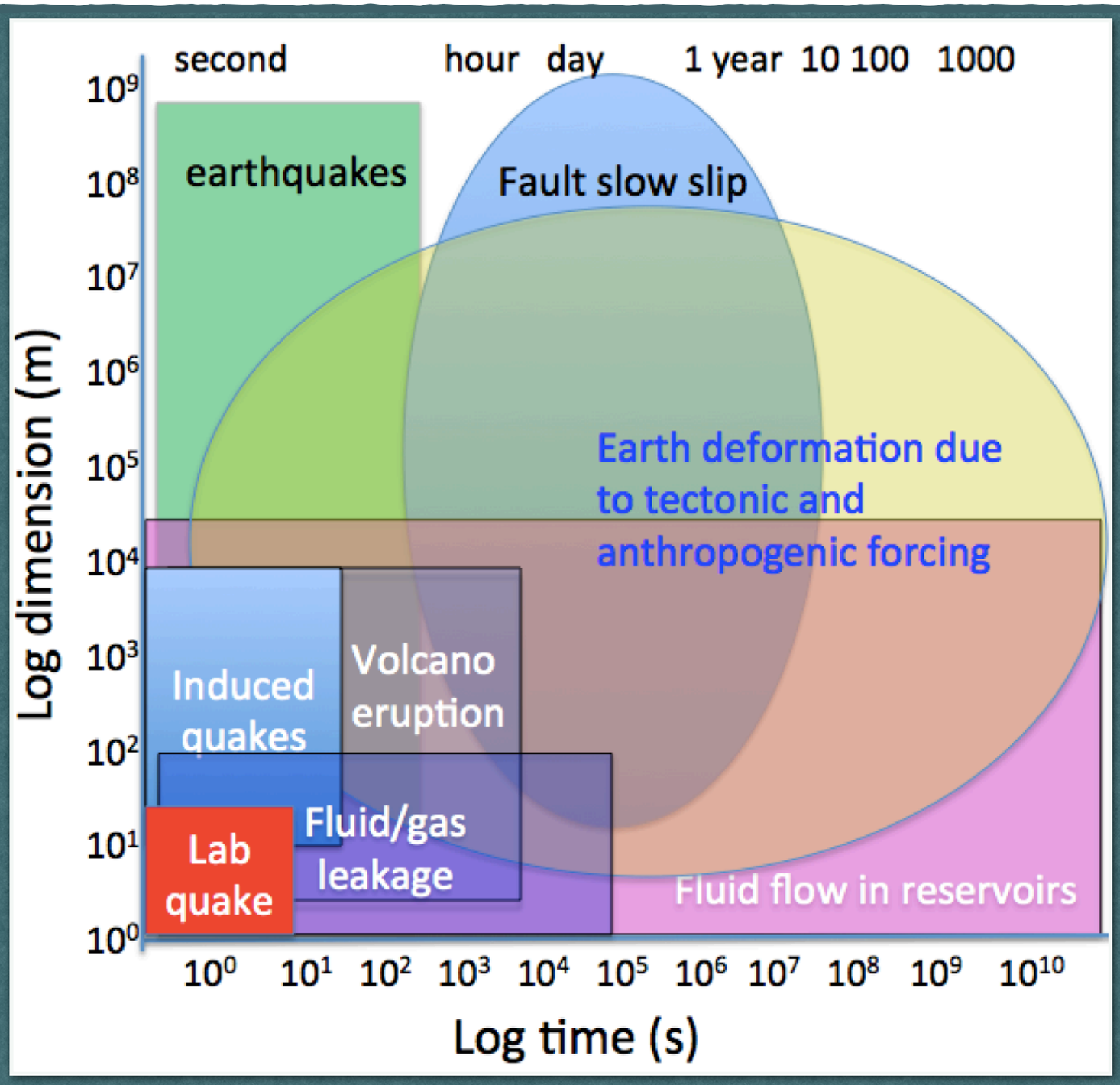
1900-1950s	**Radiometric dating**	**age of Earth**
1930's	**Magentometer**	**pole reversals**
1950s	**Nuclear testing**	**basis of modern seismology**
1960's +	**Spacecraft/satellite**	**origin of KT extinction, GPS, Earth imagery…..**
WWII	**Oceanic research vessels**	**Discovery of magnetic stripes in oceanic basalt-proof of plate tectonics (in the 1960's)**
1980's+	**Widely available computers**	**massive advances in imaging, simulating Earth processes, data processing….)**
1980's	**Invention of the WWW**	**revolutionary access to information and data**
1990's	**Energy technology advances**	**e.g., horizontal drilling**
1980'+	**Superfast Computers**	**dramatic advances in waveform inversion, large scale simulation etc.)**
2010's	**GPS, InSAR**	**dramatic advances in measuring Earth surface displacement**
present	**Gaming and GPUs**	**dramatic advances in waveform inversion, large scale simulation etc.)**

TECHNOLOGICAL SHIFTS LEADING TO REVOLUTIONARY ADVANCES IN SOLID EARTH GEOSCIENCE

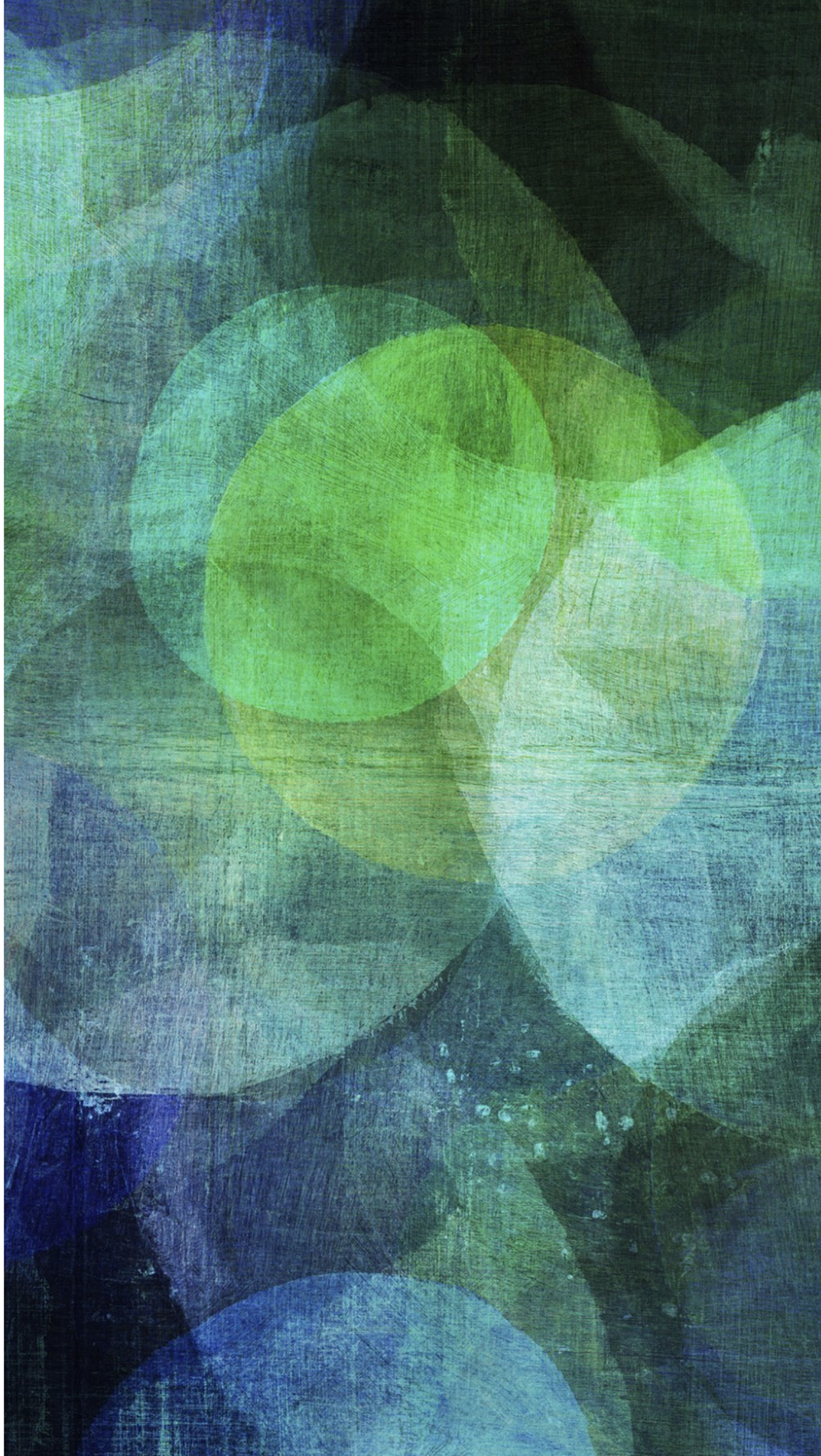
present	**Machine learning/ big data/ ultrafast computers computers**	**?**



SCALES OF GEOPHYSICAL PROBLEMS



Machine learning: some background



AI: WHY NOW?

Confluence of:

Big Data

Big Computers

Deep Architectures

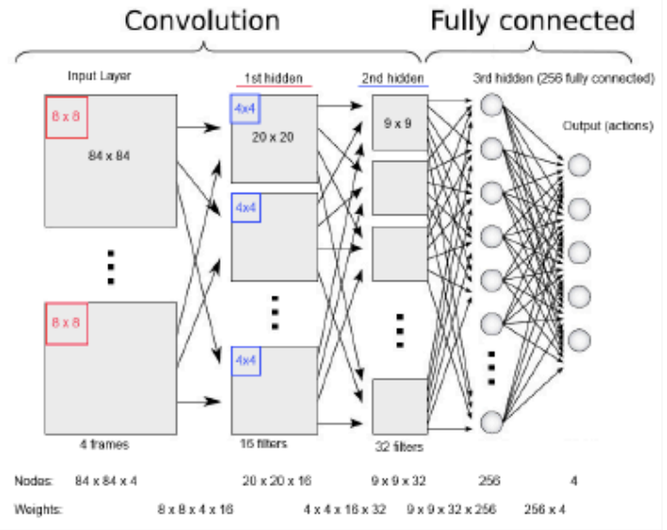


Figure courtesy Rich Baraniuk

ARTIFICIAL INTELLIGENCE

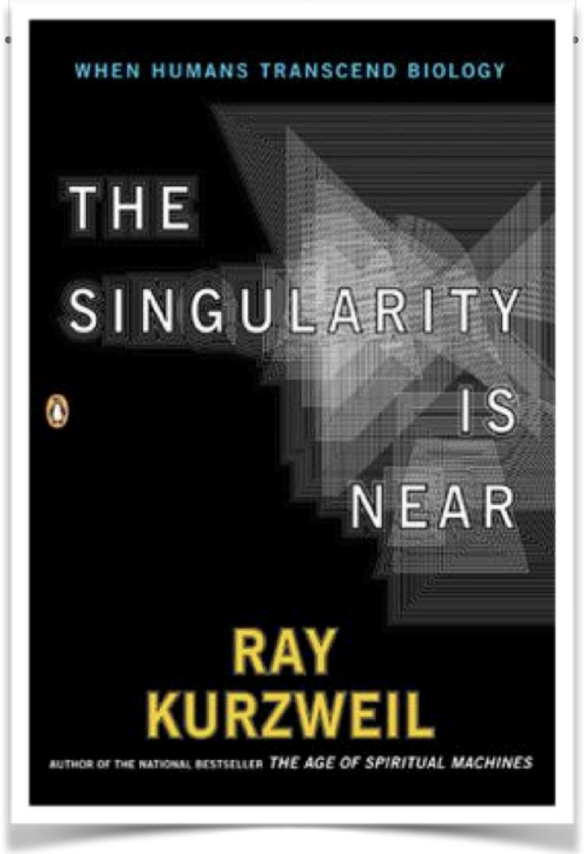
Much hyperbole, much promise

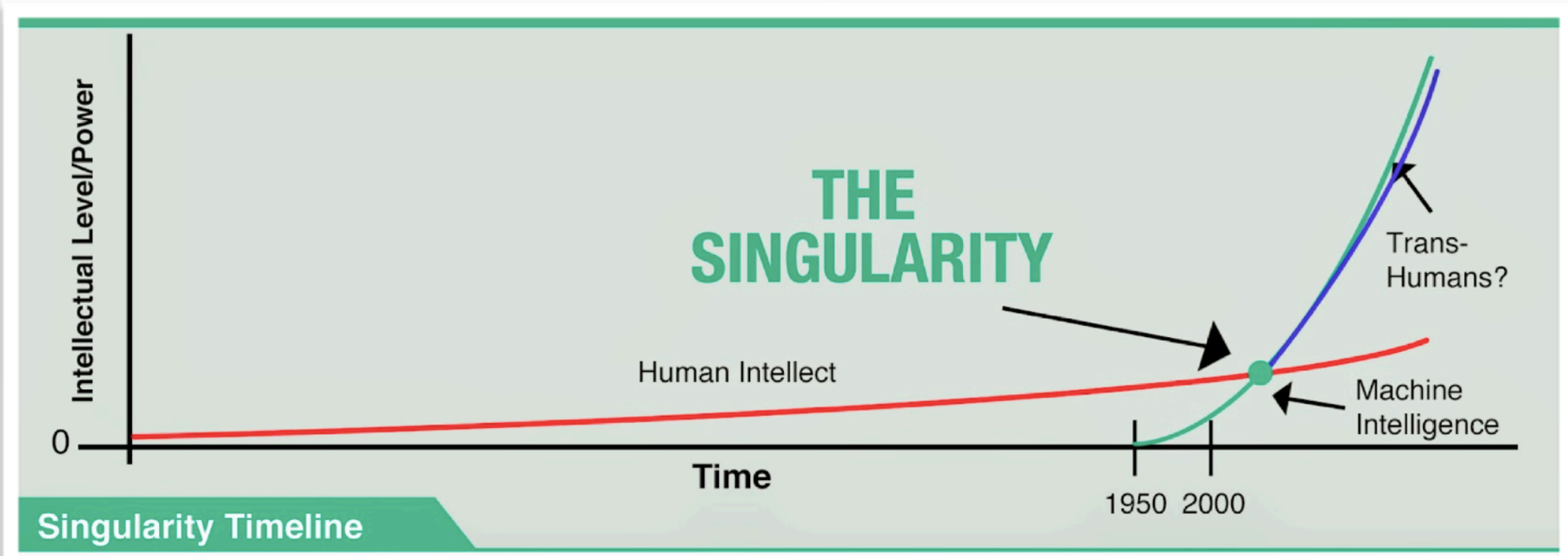
In two decades

- **data availability increased by 1,000-fold, key algorithms have improved 10-fold to 100-fold, and**
- **hardware speed has improved by at least 100-fold.**

Ninety percent of the digital data in the world today has been created in the past two years alone.

THE SINGULARITY





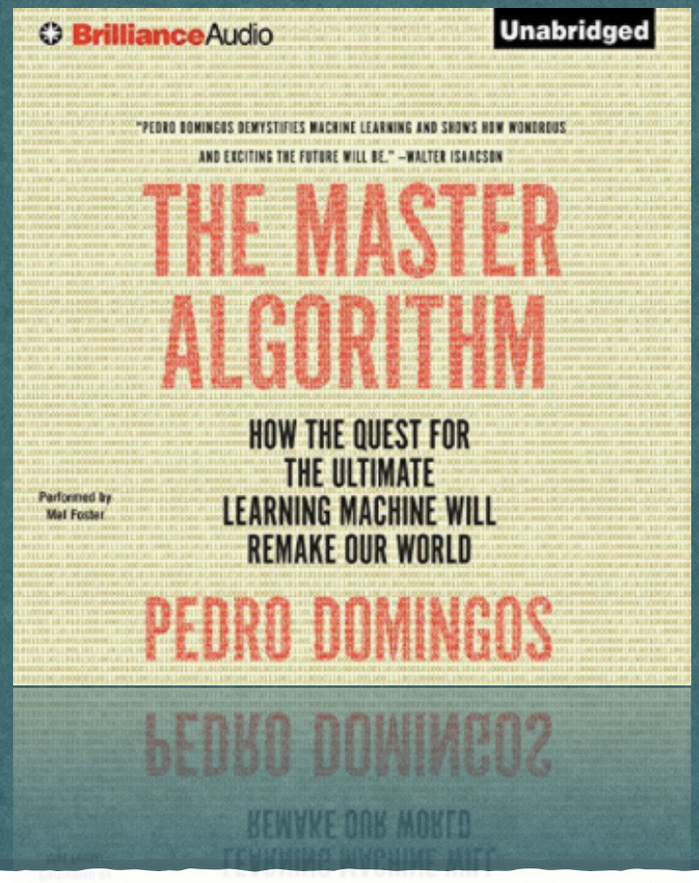
Singularity Timeline

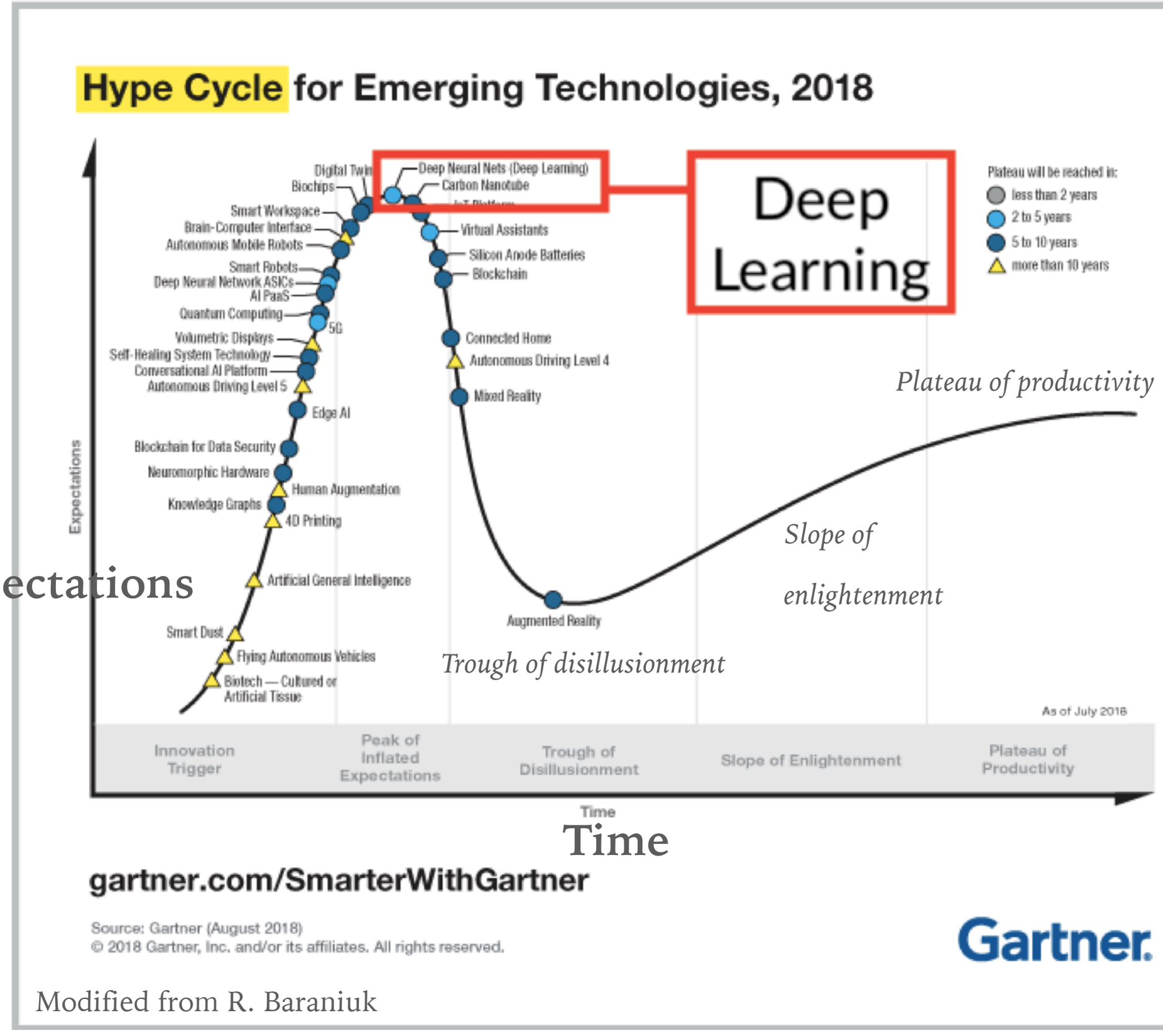
Rise in human intellect could be driven by integrating with machines in the future

AI

"People worry that computers will get too smart and take over the world,

but the real problem is that they're too stupid and they've already taken over the world."



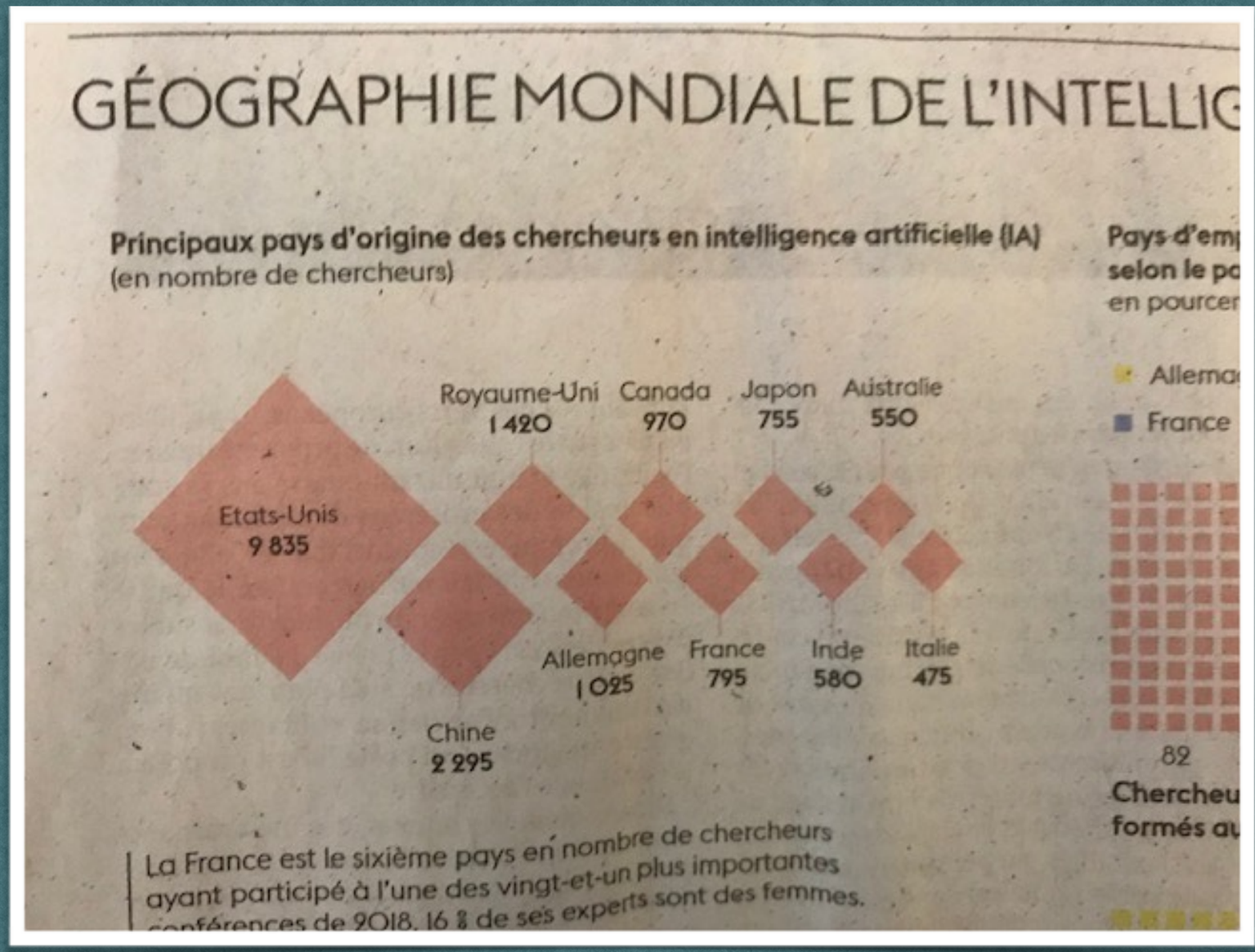


Modified from R. Baraniuk

AI RESEARCHERS PER COUNTRY

La Monde, June 5, 2019

Le Monde


GÉOGRAPHIE MONDIALE DE L'INTELLIG

Principaux pays d'origine des chercheurs en intelligence artificielle (IA)
(en nombre de chercheurs)

Etats-Unis 9 835
Chine 2 295
Royaume-Uni 1 420
Canada 970
Allemagne 1 025
France 795
Japon 755
Inde 580
Australie 550
Italie 475

La France est le sixième pays en nombre de chercheurs ayant participé à l'une des vingt-et-un plus importantes conférences de 2018. 16 % de ses experts sont des femmes.

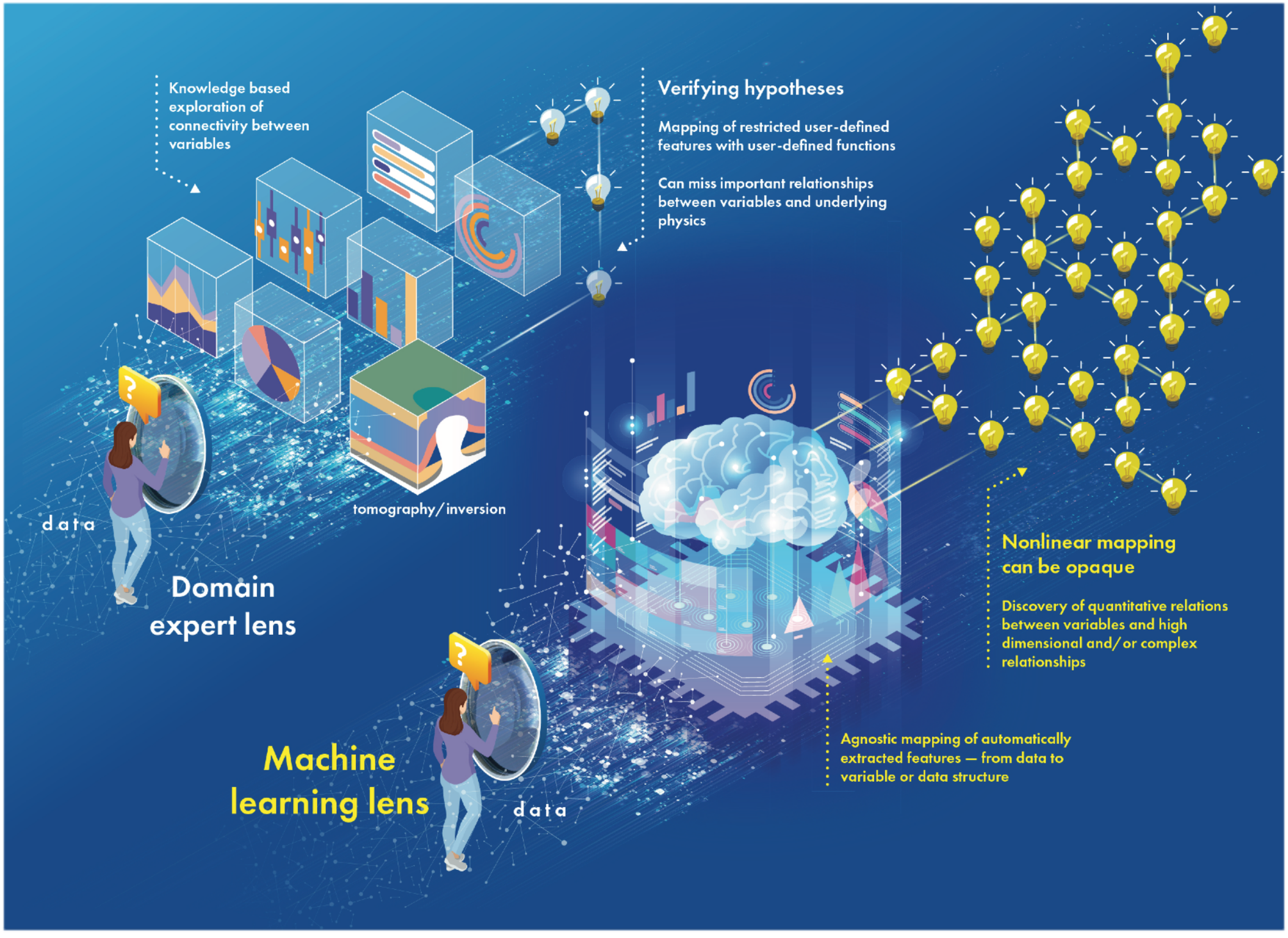
Pays d'emp
selon le po
en pourcer

Allema
France

82

Chercheu
formés au

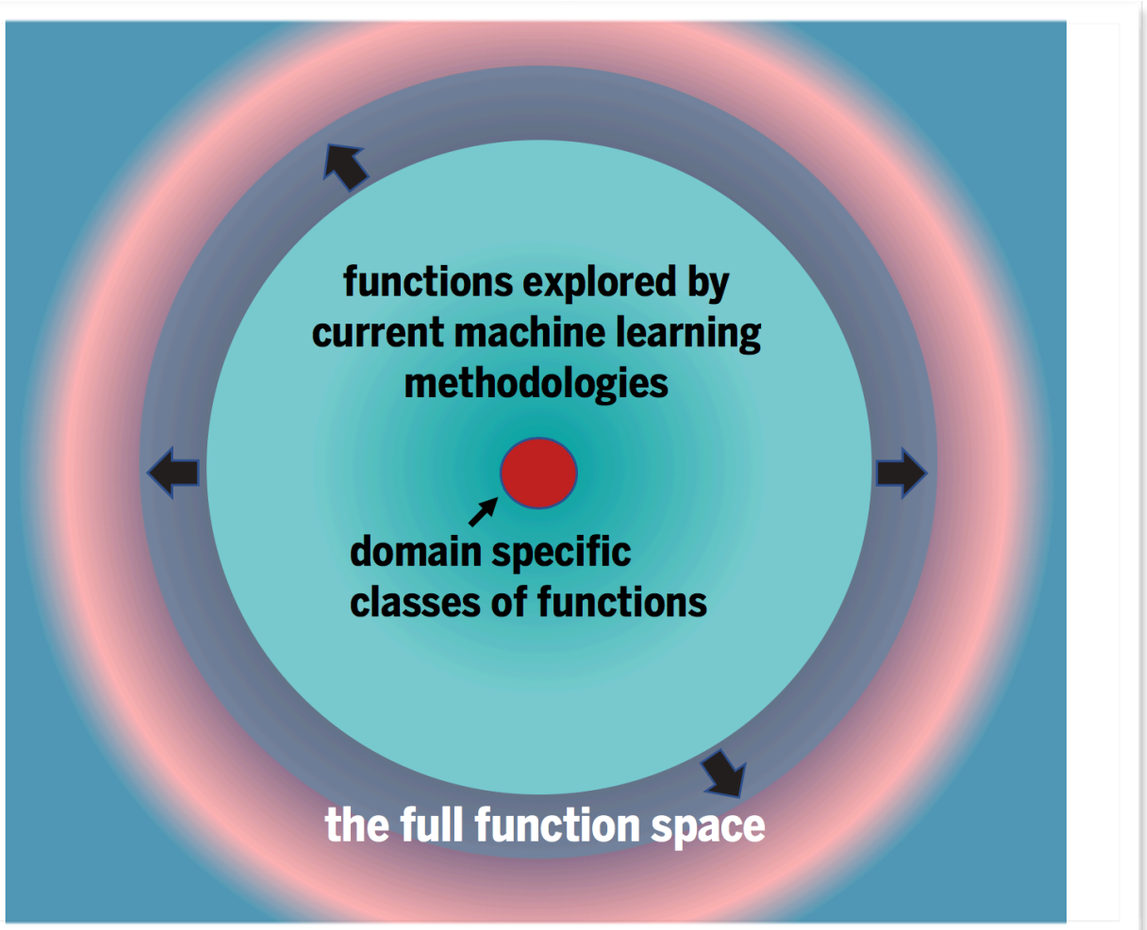
THE MACHINE LEARNING LENS



Bergman, Johnson, de Hoop, Beroza *Science* (2019)

WHAT DOES IT MEAN TO APPLY THESE APPROACHES TO THE SUBSURFACE?

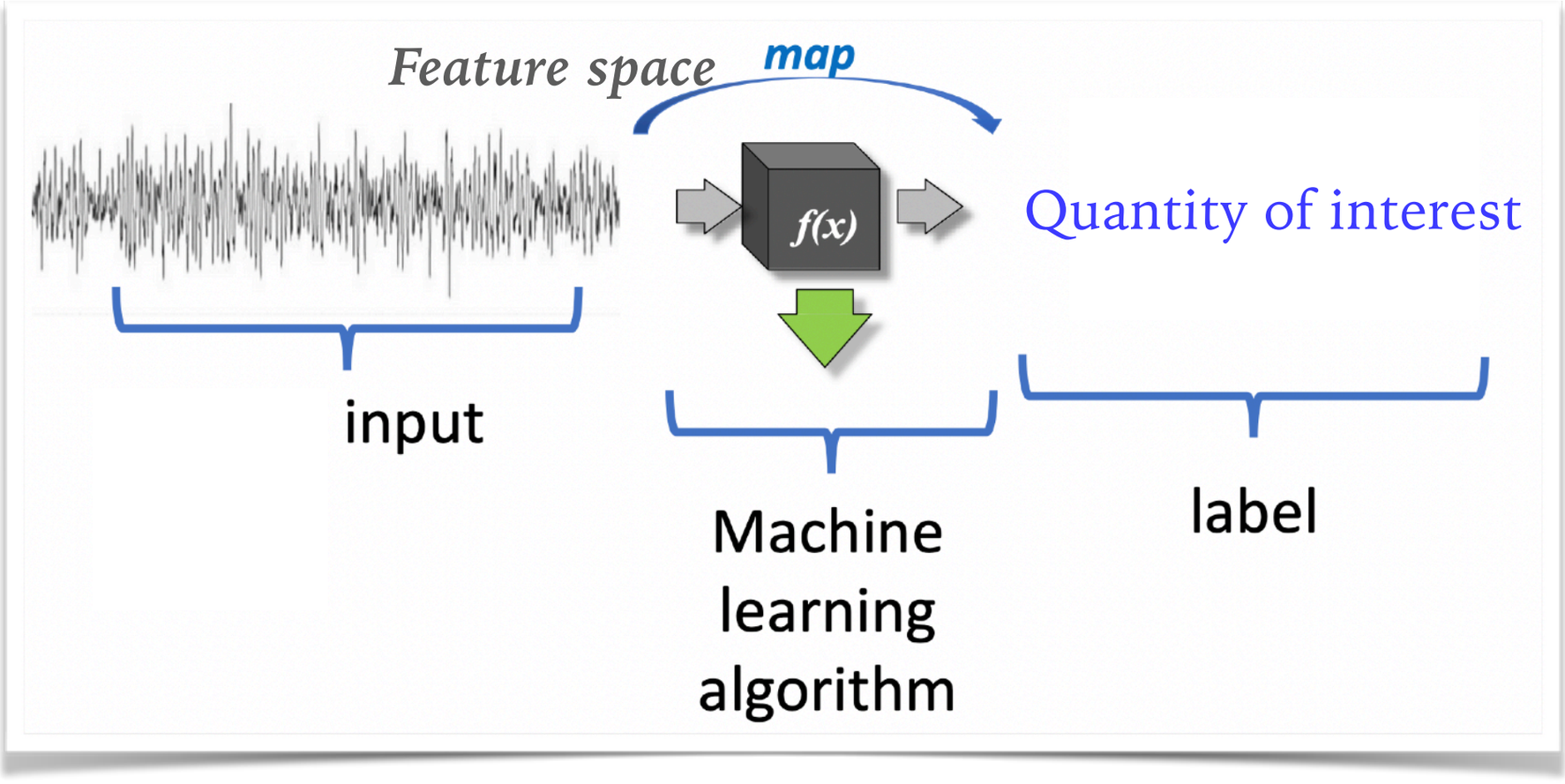
- **ML/data analytics are a new tool in our toolbox.**
  Simply a large group of functions simultaneously applied to data
  Enormous advantage because of the function space explored



Bergen, Johnson, de Hoop and Beroza, *Science* (2019)

SUPERVISED LEARNING

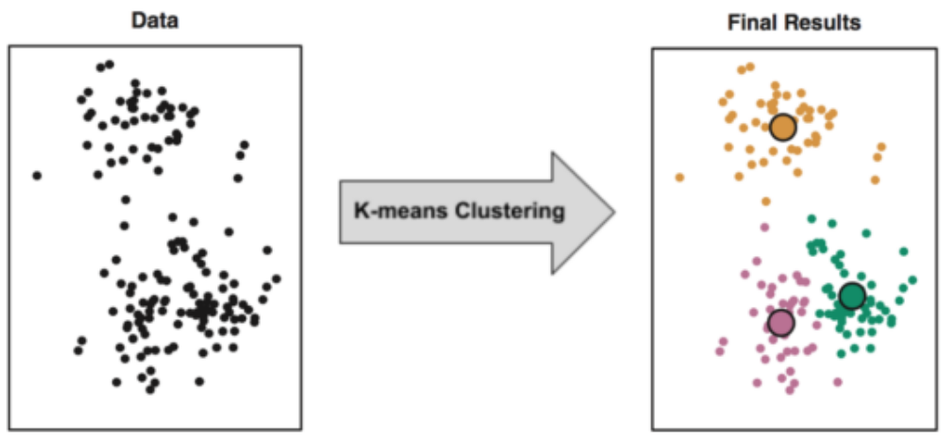
Machine learning generally involves a training procedure to build the algorithm, and then testing on a new set of data



UNSUPERVISED LEARNING

- **Unsupervised learning**: computers learn to "teach themselves"
  - model underlying structure or distribution in the data
  - there is **no correct answer and there is no teacher.**
  - algorithms left to discover interesting structure in the data

Simple example: classification, e.g. clustering, association k-means for clustering problems.



UNSUPERVISED AND SUPERVISED LEARNING: DEEP NEURAL NETWORKS

WIRED STAFF SCIENCE 06.26.12 11:15 AM

GOOGLE'S ARTIFICIAL BRAIN LEARNS TO FIND CAT VIDEOS

By Liat Clark, Wired UK

When computer scientists at Google's mysterious X lab built a neural network of 16,000 computer processors with one billion connections and let it browse YouTube, it did what many web users might do – it began to look for cats.

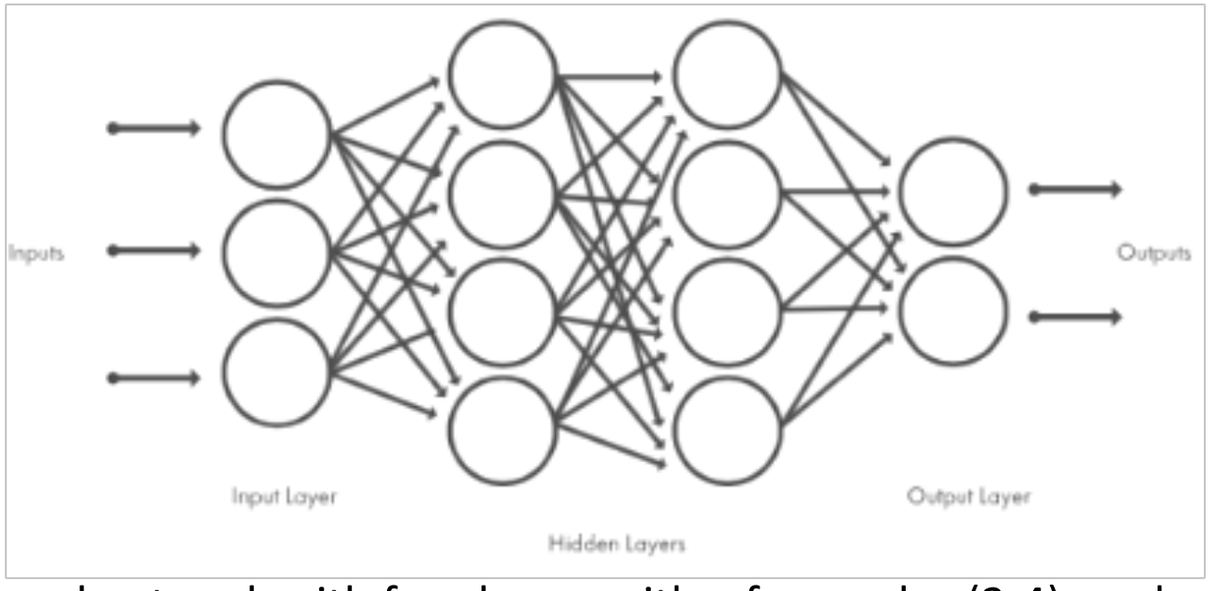
"We never told it during the training, 'This is a cat,'" Jeff Dean, the Google fellow who led the study. "It basically invented the concept of a cat."

"The idea is that you throw a ton of data at the algorithm and have the software automatically learn from the data," -- Andrew Ng, Stanford University.

DEEP LEARNING (UNSUPERVISED AND SUPERVISED APPLICATIONS)

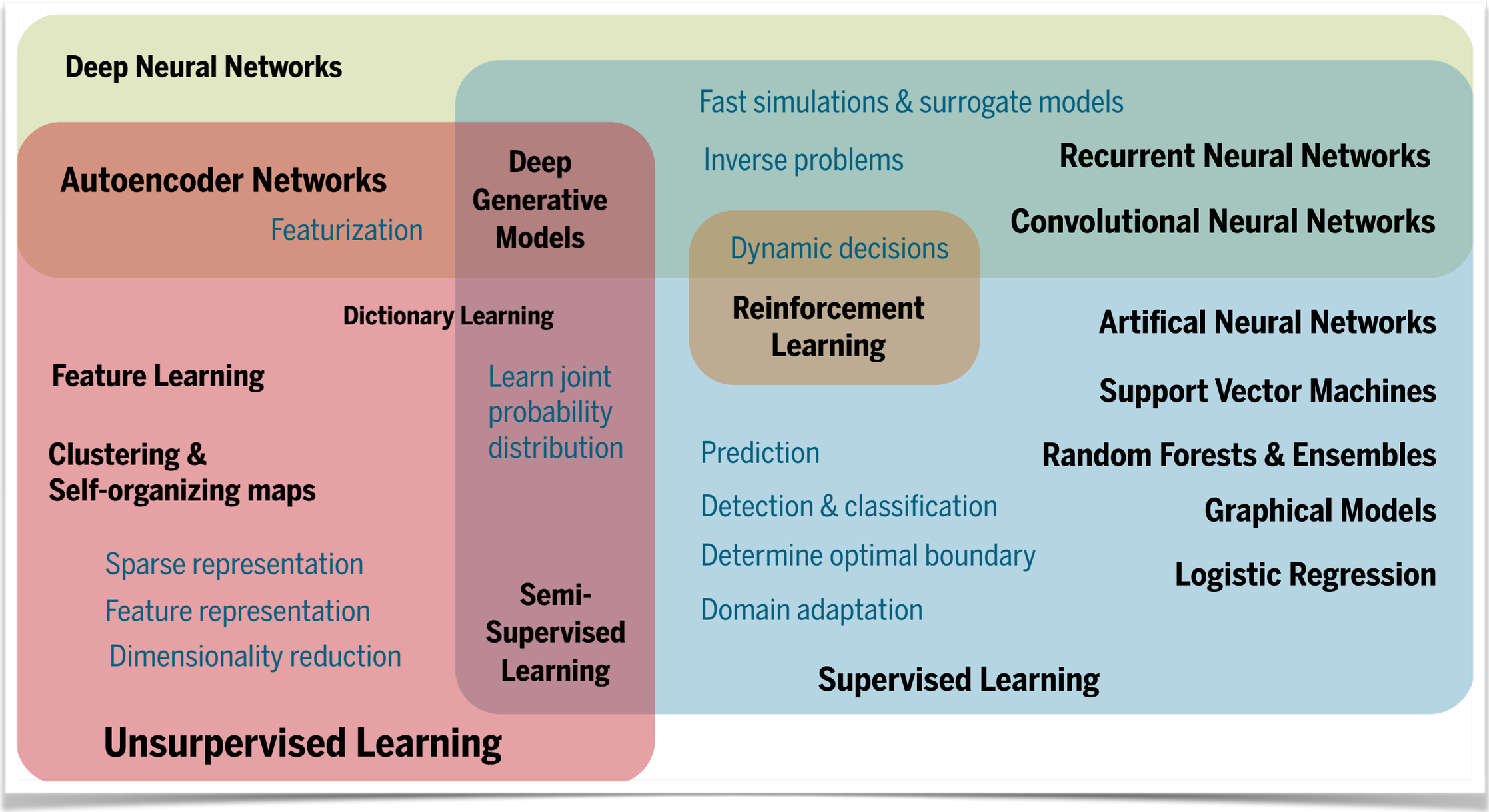
- much more powerful than traditional ML algorithms.
- **Deep neural networks are great at feature extraction: the process of figuring out what aspects of a dataset are actually useful for making predictions.**

The word ‘deep’ comes from the fact that AI developers usually use networks with tens or hundreds of layers (and tens or hundreds of nodes per layer).



e.g., Neural network with four layers with a few nodes (2-4) per layer.

MACHINE LEARNING IS COMPRISED OF A ZOO OF TECHNIQUES



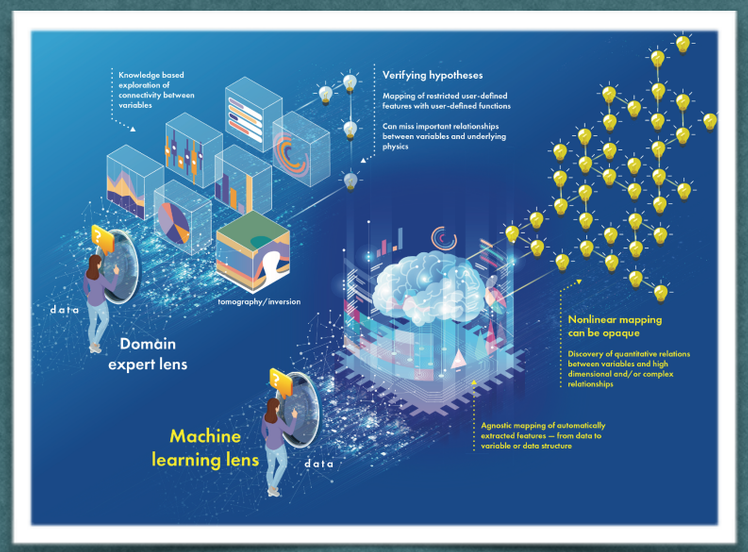
Bergen, Johnson, de Hoop and Beroza, *Science* (2019)

DATA AND APPROACHES

- Most data appropriate
- Data **quality** is key
- '**labelled' data** for supervised learning fundamental
- ML appropriate when unknown, unexpected function space must be explored

When are other approaches appropriate?

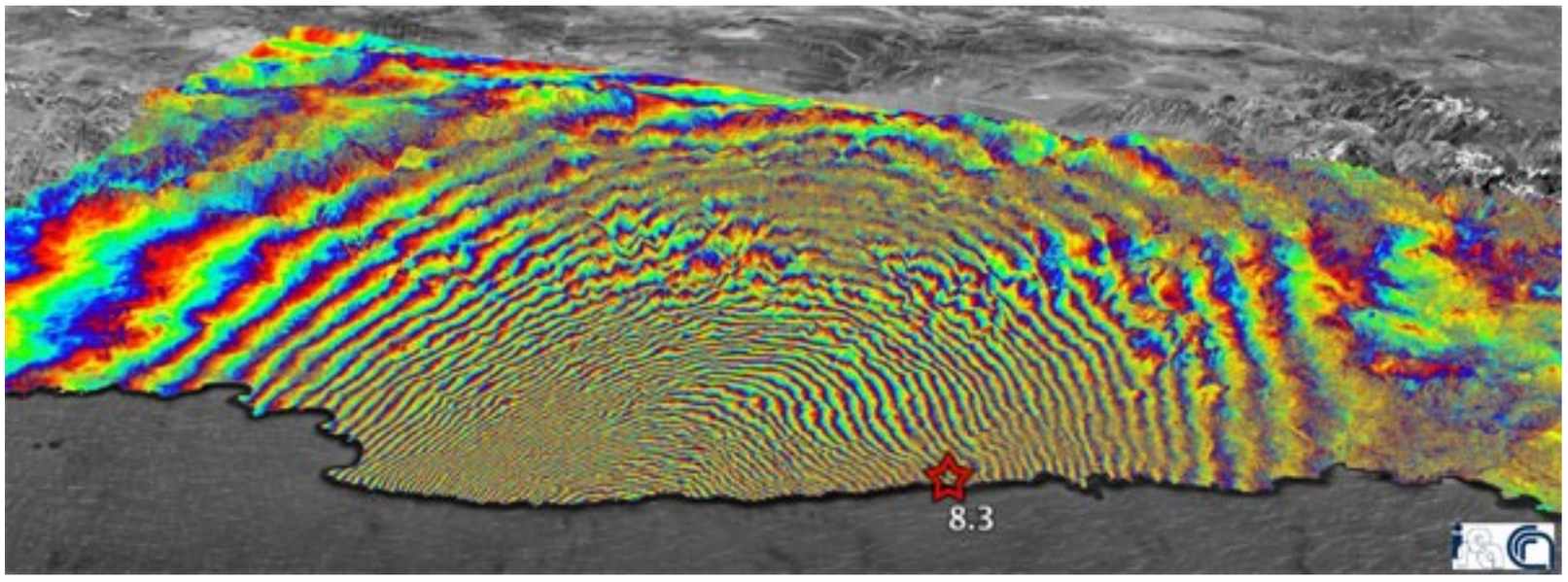
Simple questions may require simple approaches:
e.g., is an FFT sufficient?



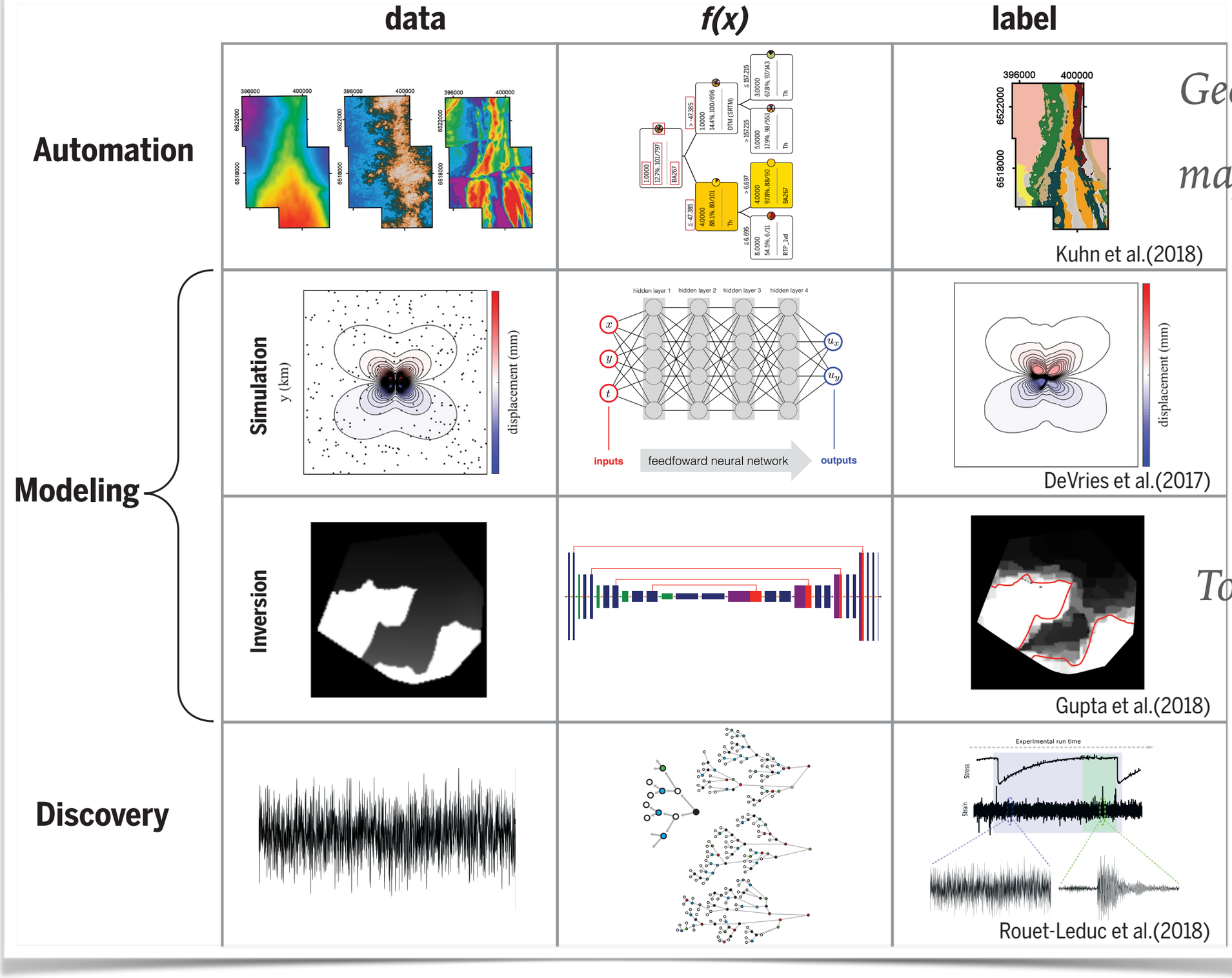
WHAT TOPICS CAN ML/AI HELP ANSWER IN GEOSCIENCE?

Nearly every topic in geoscience has applications!

- InSAR
- GPS
- Seismic Imaging (industry and basic research)
- Earthquake catalogs
- Gravity
- EM
- Geology
- Fluid flow in porous media
- Continuous geophysical analysis of all kinds…..



EXAMPLES OF GEO-APPLICATIONS



Bergen, Johnson, de Hoop and Beroza, *Science* (2019)

WHEN DOES MACHINE LEARNING FAIL?

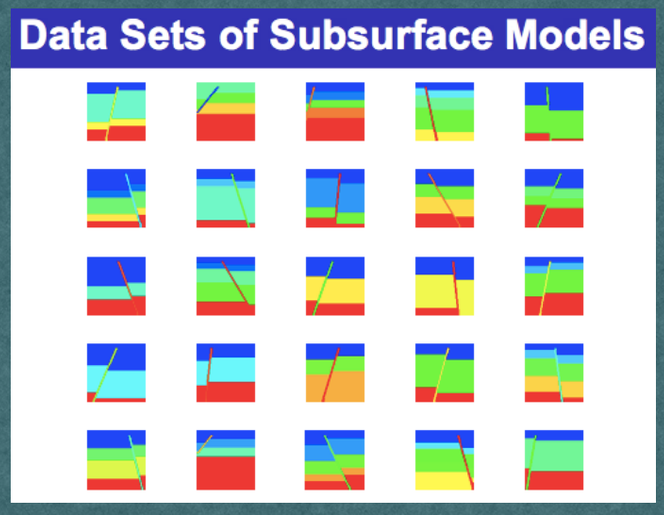
1. **nuisance variation in images**: changes in location, pose, viewpoint, lighting, expression, occlusion, such as Landsat imagery—
2. **Non-stationary data**
3. **Randomness** or **Enthropy**.: You cannot learn a pattern that does not exist
4. **A lack of training data** is one of the most common reasons why machine learning can fail.
5. **over-fitting:** classifier only recognizes what it has seen because the distribution of your training data does not capture the true distribution of the pattern you want to learn [Marko Živković,].

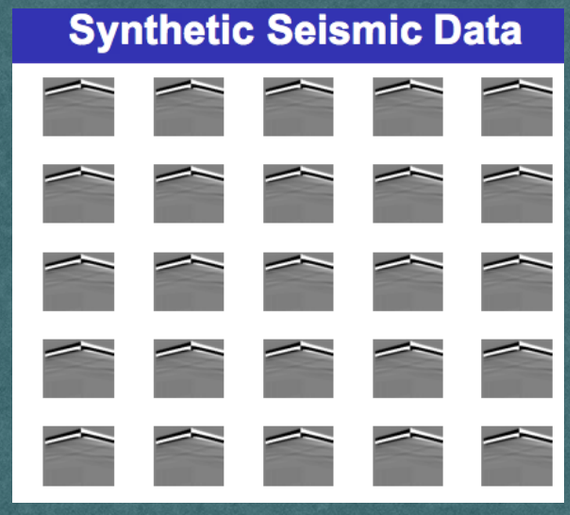
Work with experts—domain expert + ML expert a very good combination

Other challenges:
Obtaining data….

GOING TO THE FUTURE

- Geo Workshops 2019 (Montreal) and 2020 (Vancouver)
  Neurological Information Processing Systems (NeurIPS)
- Special sessions AGU, Seism. Soc. Am. 2018, 2019
- Special sessions SEG, IEEE
- Industry focused meetings





GOING TO THE FUTURE

WHEN

Monday, March 18, 2019 - Friday, March 22, 2019

Add to Calendar

WHERE

Santa Fe, New Mexico
USA

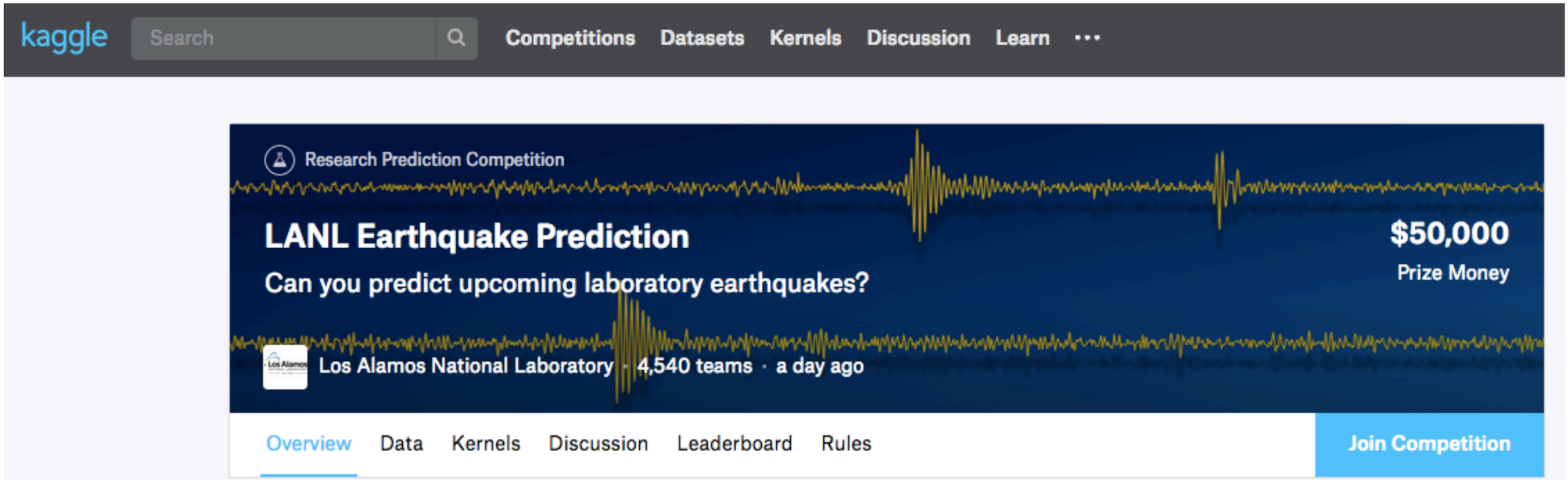
Registration opening soon!

SUMMARY

Machine learning (ML) in its current form is relatively new to geoscience. In the past, ML was applied to a number of geoscience problems but the number of applications before about 2010 was modest. These applications of ML did not reach their full potential for three primary reasons: scarcity of sufficient data for training and testing, the

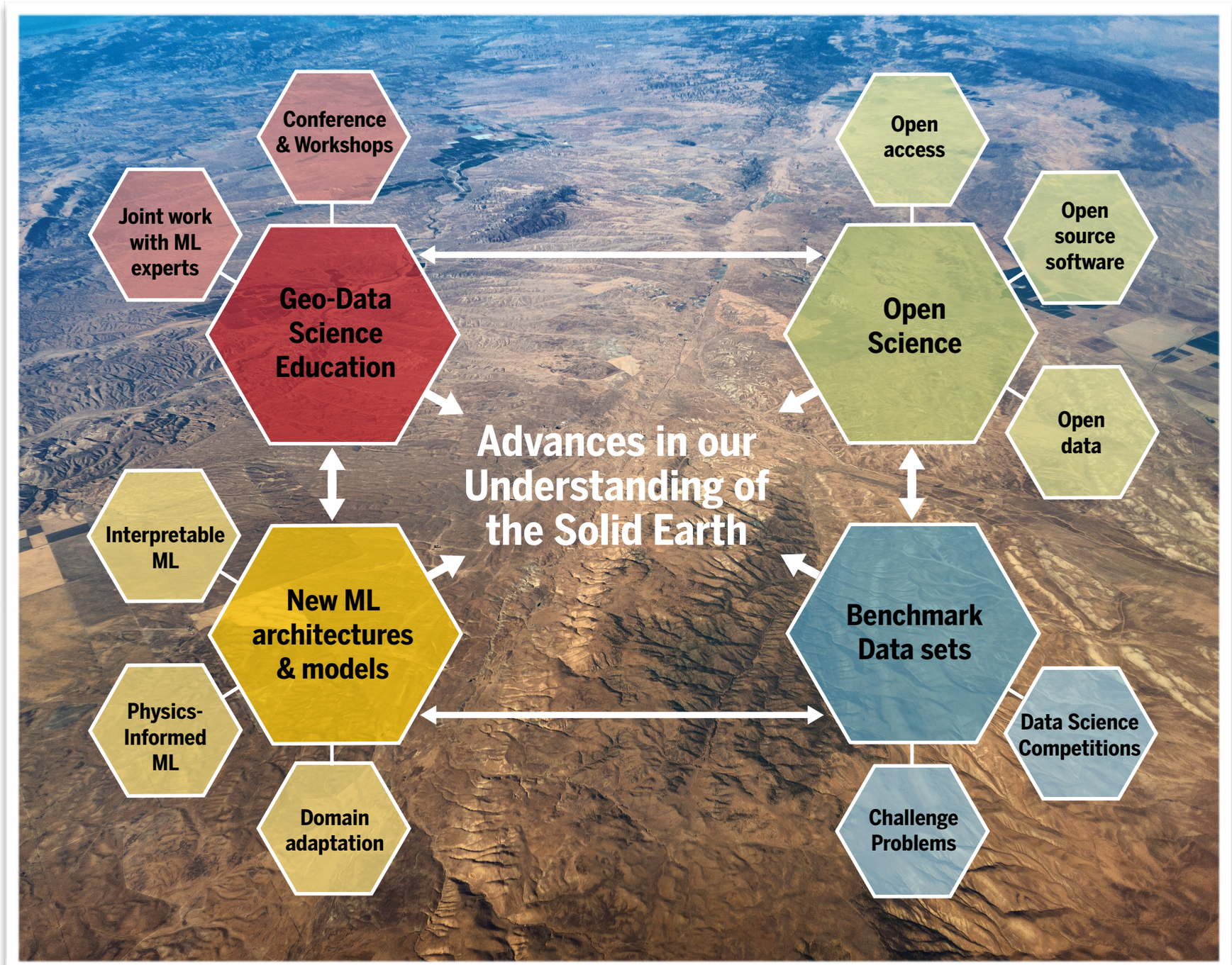
2020 meeting March in Santa Fe

COMPETITIONS



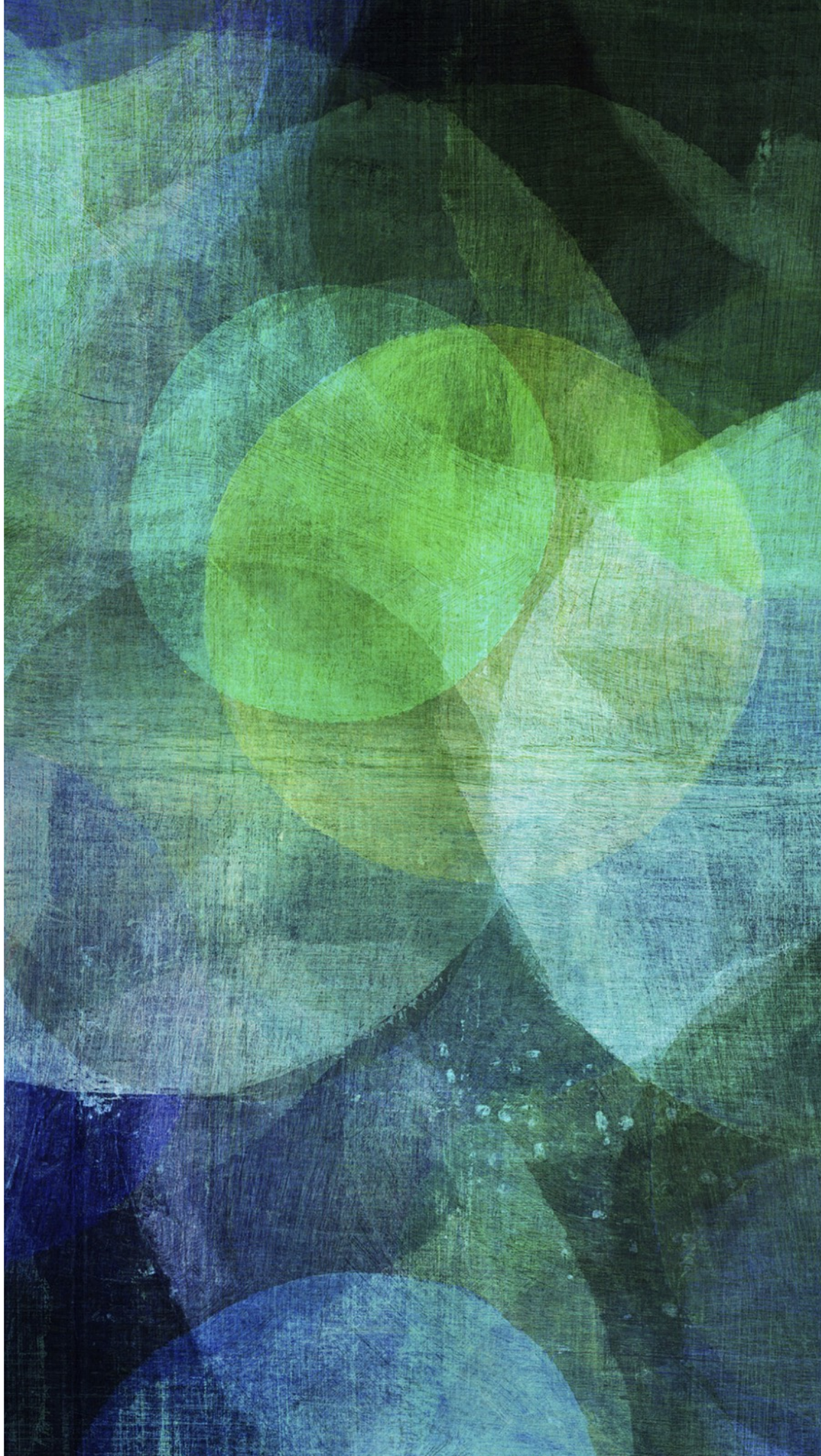
U.S. DEPARTMENT OF ENERGY | Office of Science

CONCLUSION: HOW DO WE PROCEED?



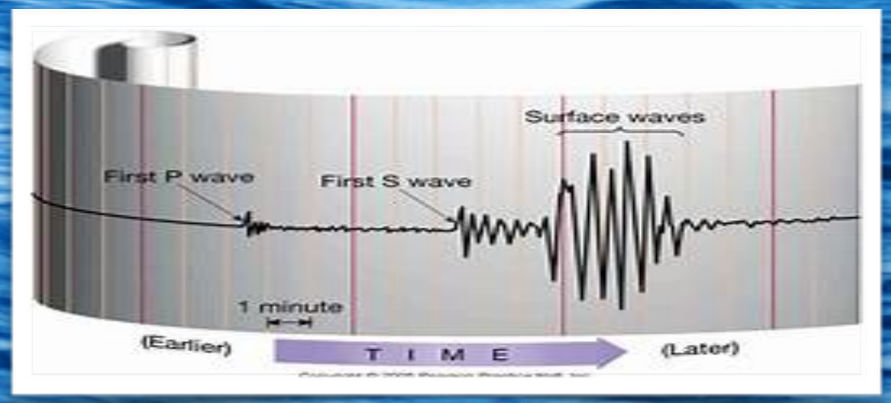
BACKUP SLIDES ON SOME OF THE LOS ALAMOS WORK

Geophysical studies of faults applying ML: a supervised learning approach



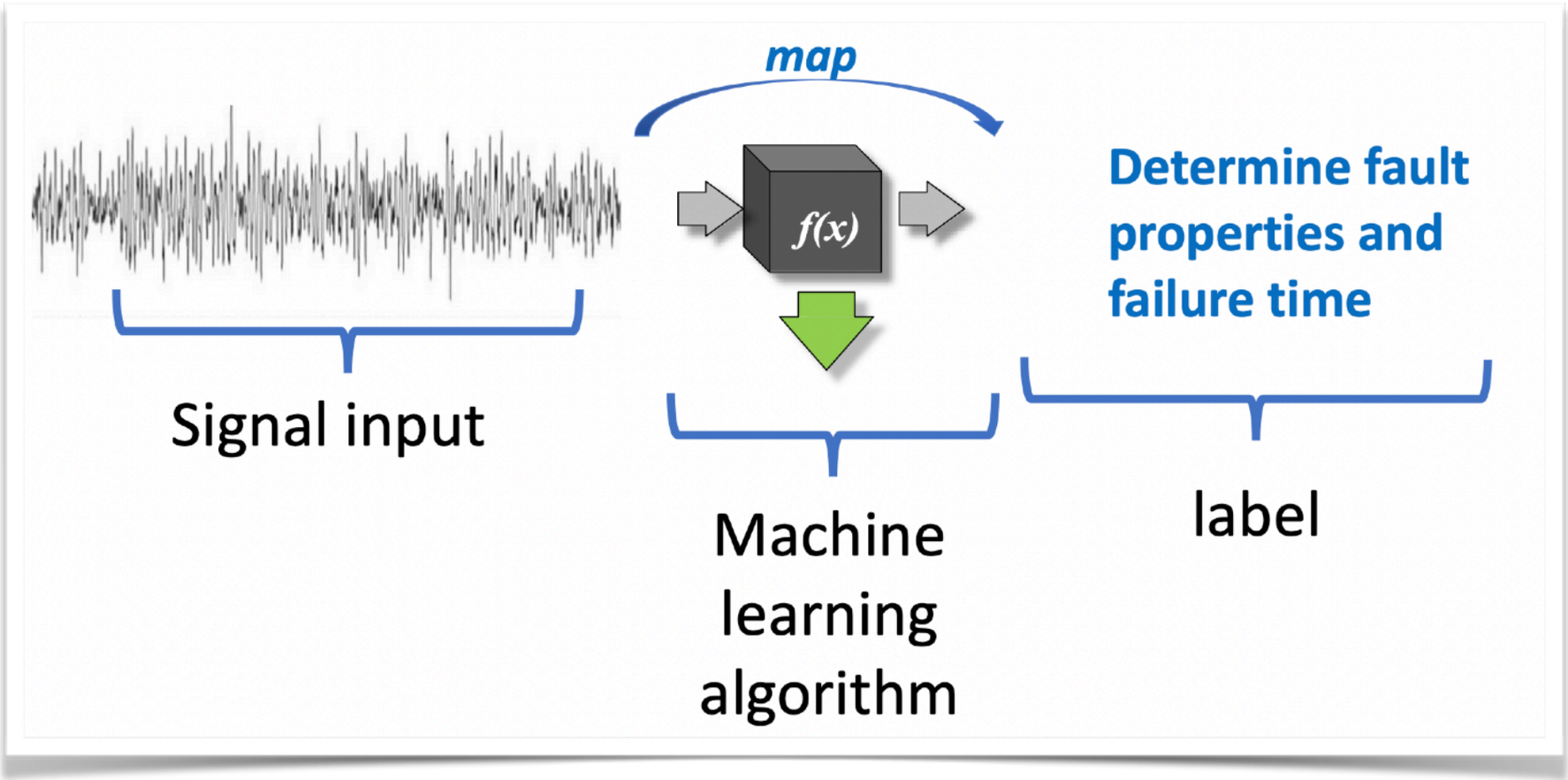
SEISMIC SIGNALS

Slip on earthquakes is manifest by individual seismic signals in a sea of background noise that tell us an earthquake took place.

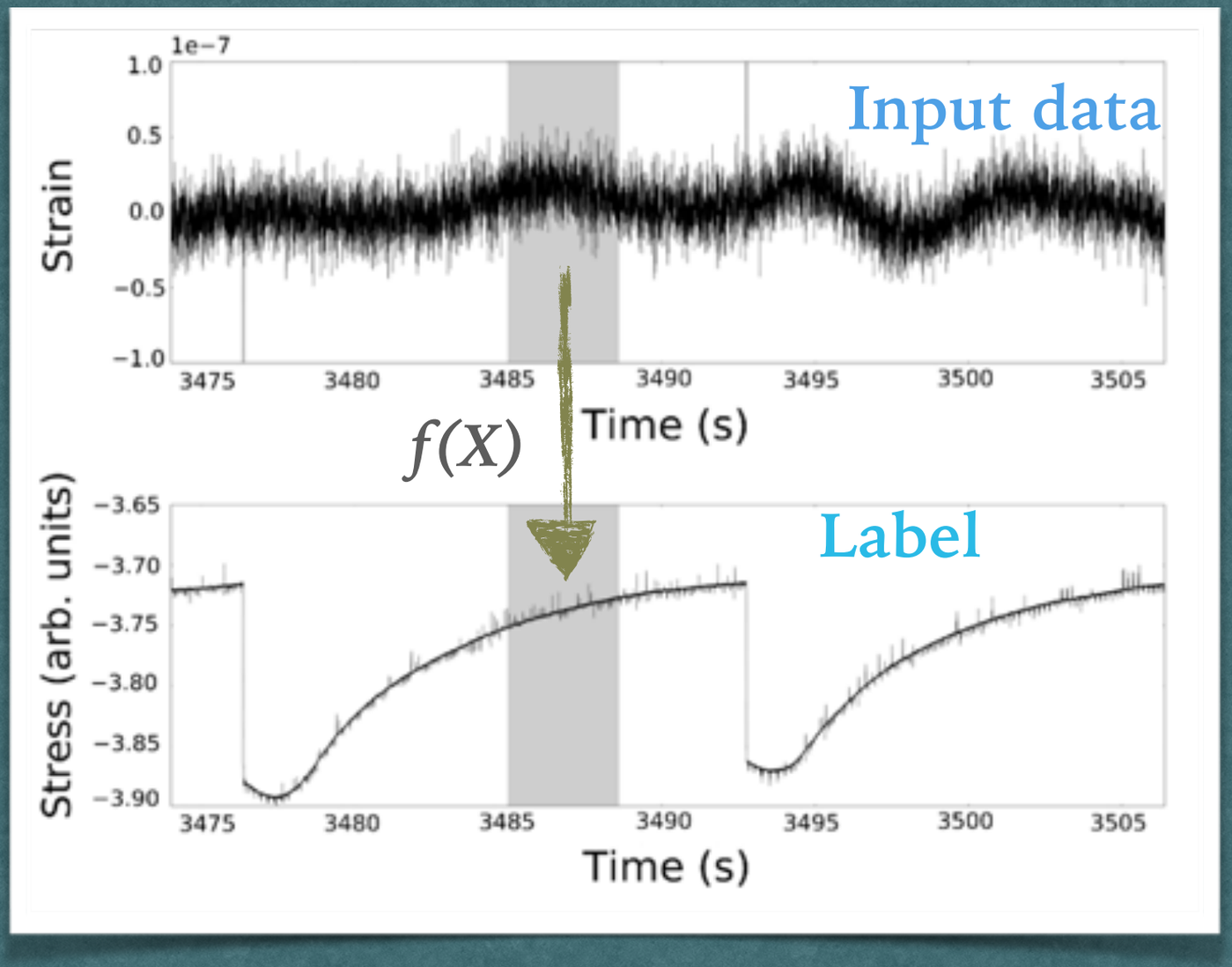


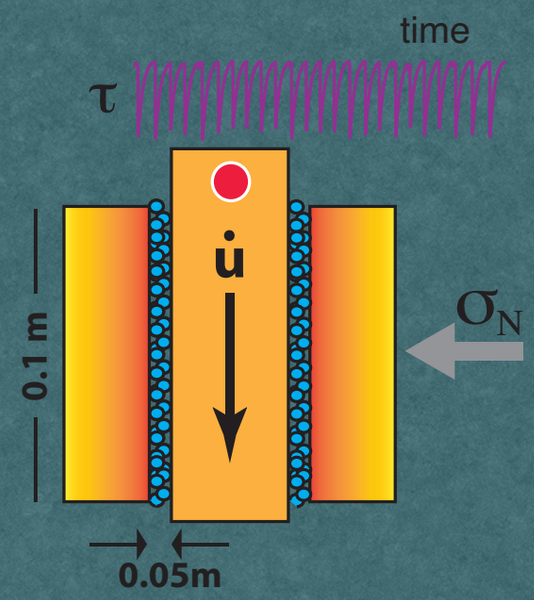
SUPERVISED LEARNING

Machine learning generally involves a training procedure to build the algorithm, and then testing on a new set of data



LABORATORY STUDY OF FAULT PHYSICS

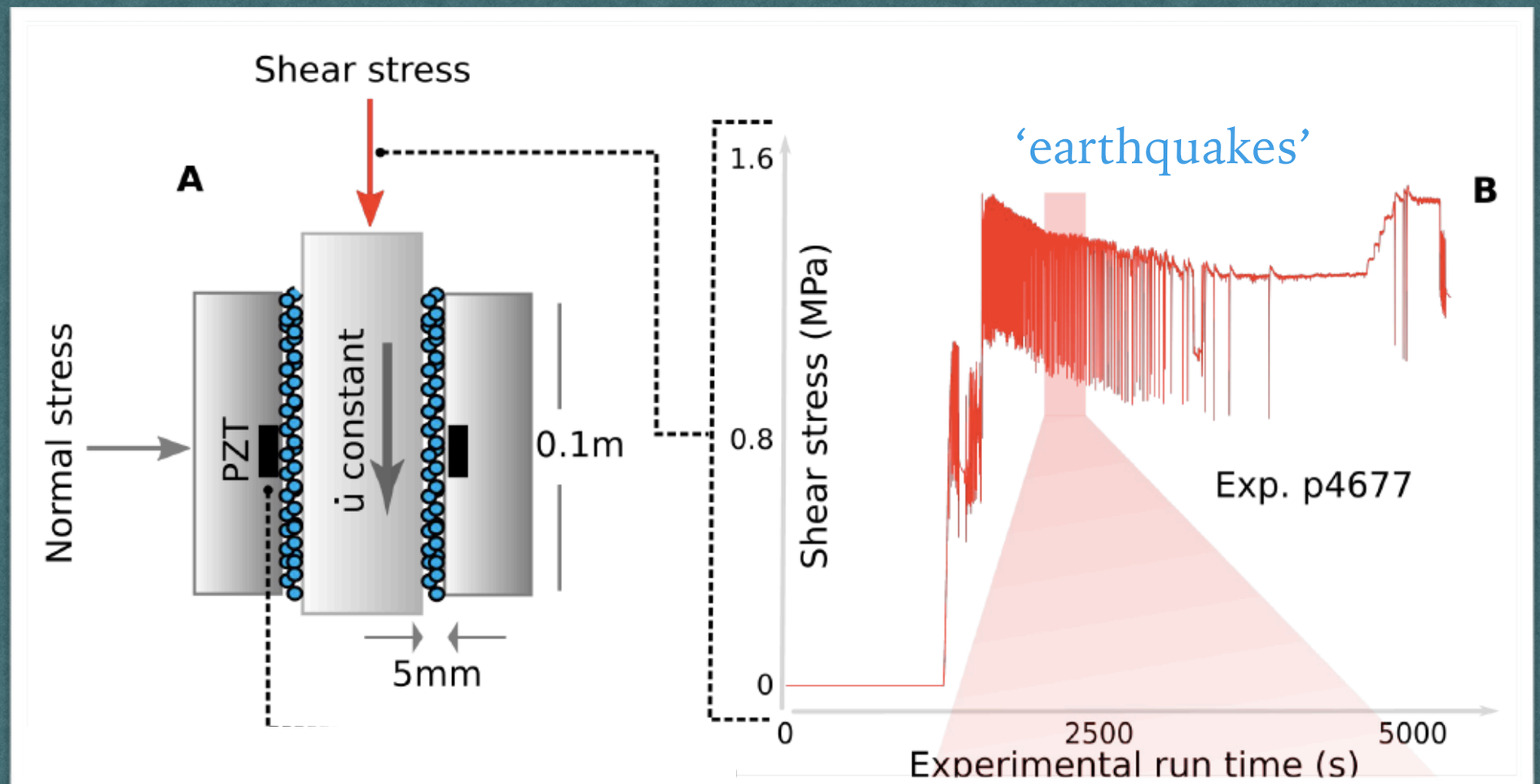




Can we learn the fault mechanics using only the acoustical signal recorded on the experiment?

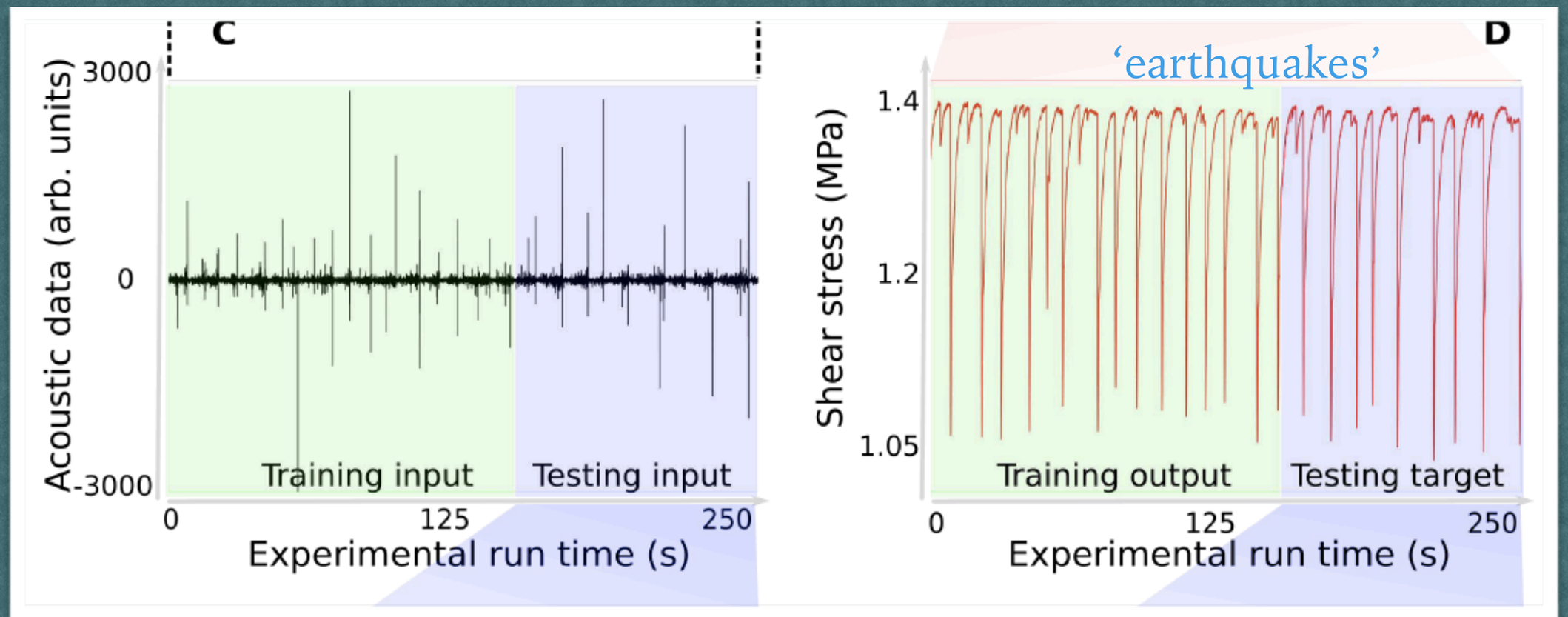
Rouet-LeDuc et al, Geophys. Res. Lett, 2017; 2018

ML RESULT



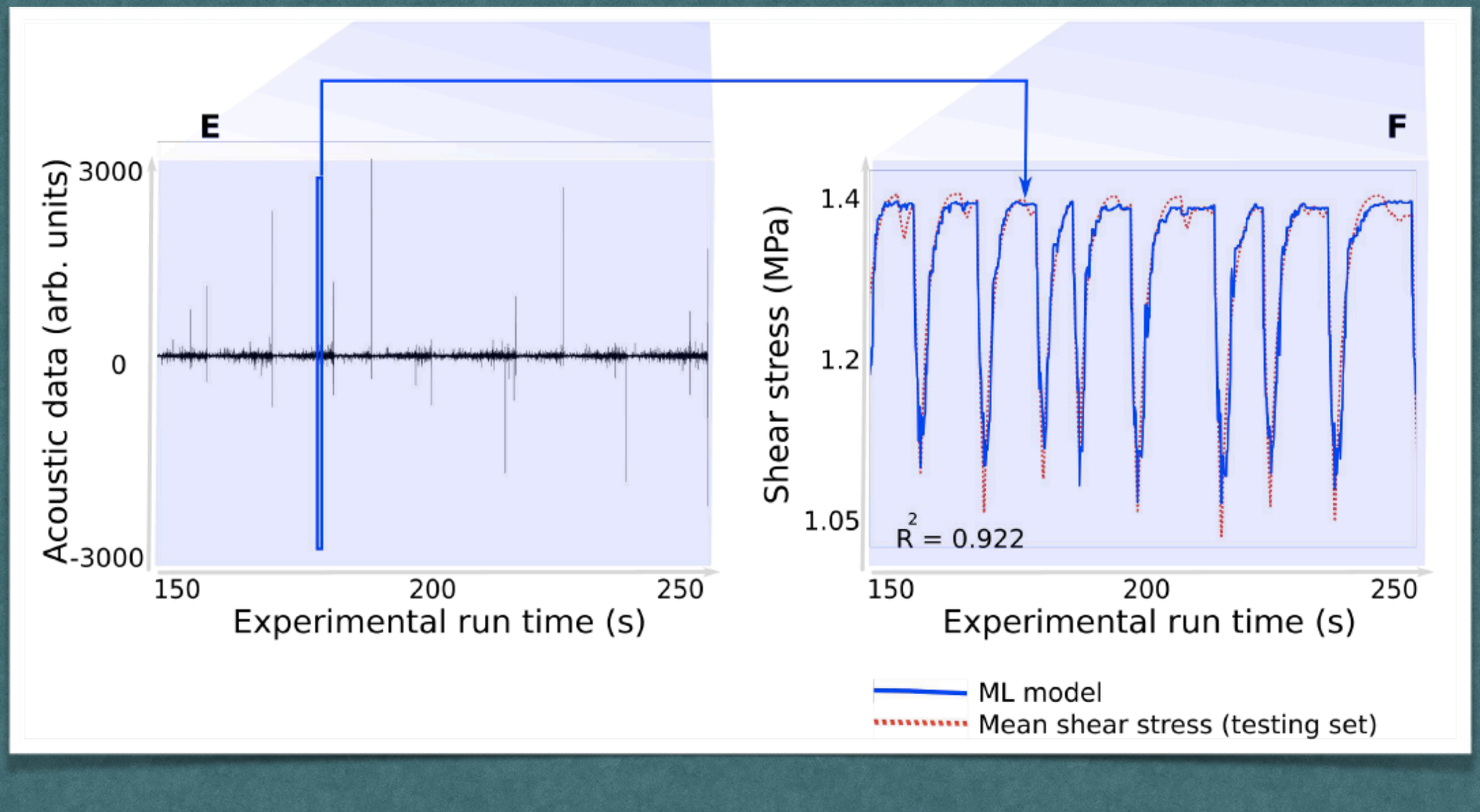
Rouet-LeDuc et al., Geophys. Res. Lett, 2018

ML TRAINING AND TESTING DATA SETS



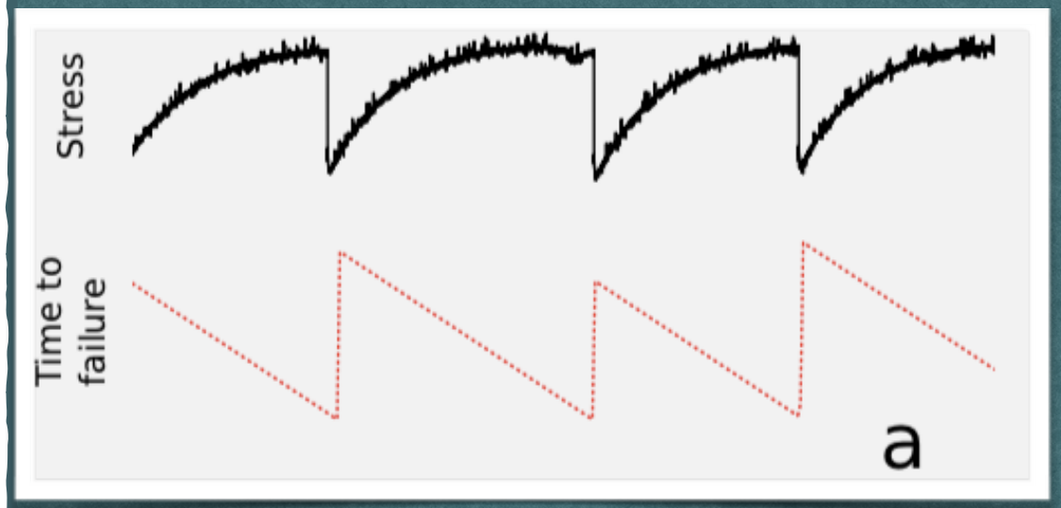
Rouet-LeDuc et al., Geophys. Res. Lett, 2018

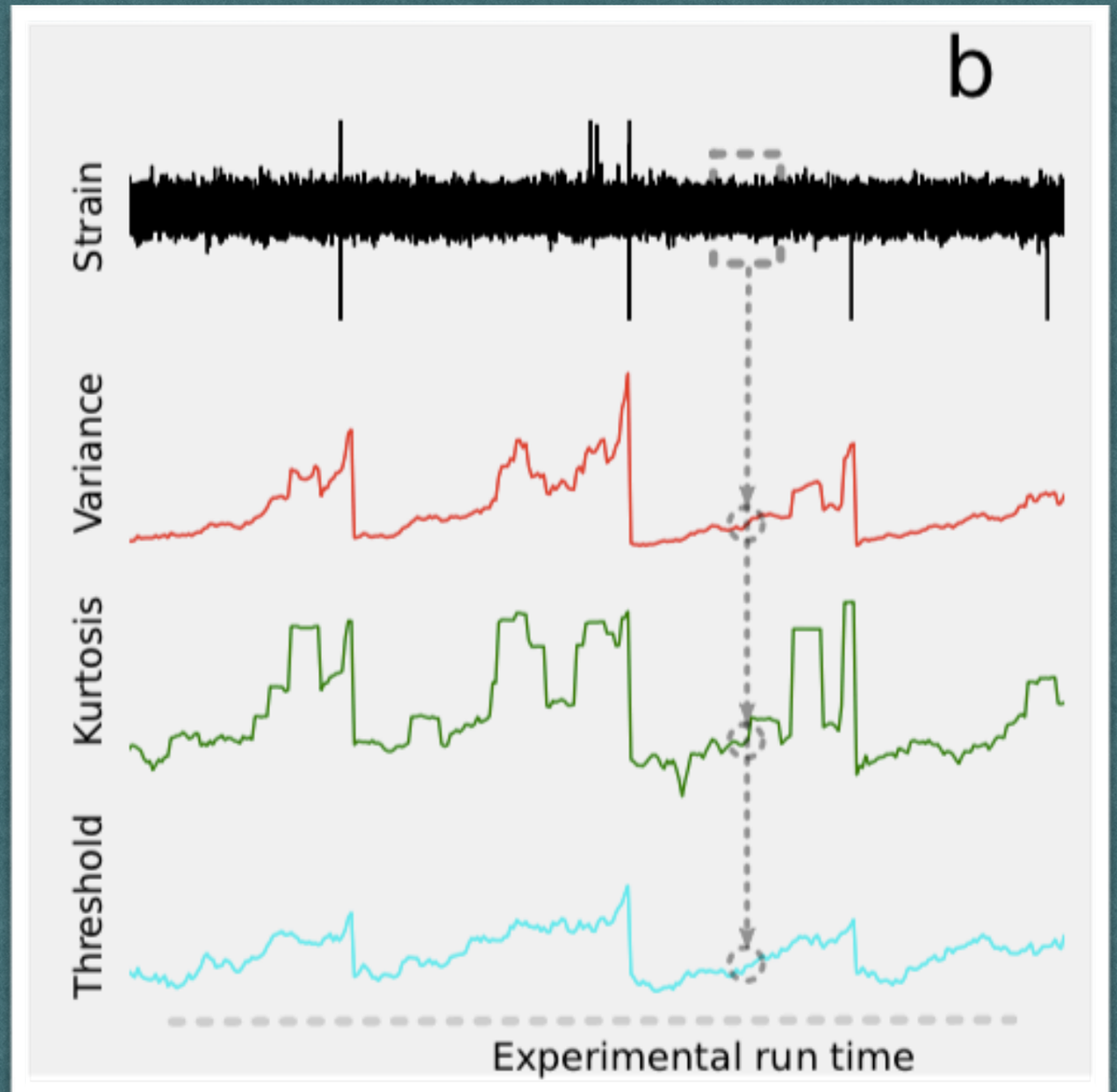
TESTING: ML RESULT



Rouet-LeDuc et al., Geophys. Res. Lett, 2018

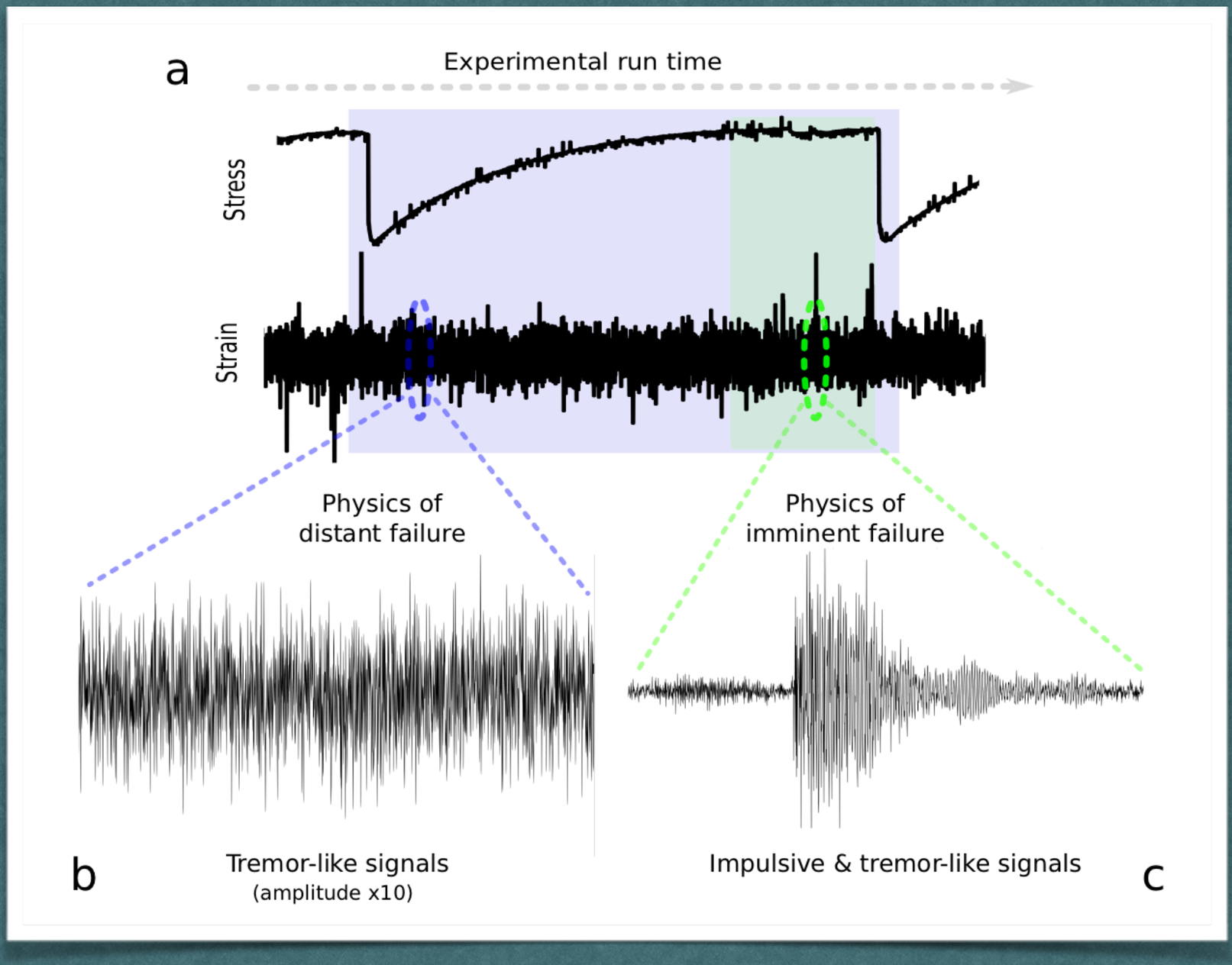
EVENT TIMING FORECAST





Rouet-LeDuc et al., Geophys. Res. Lett, 2017

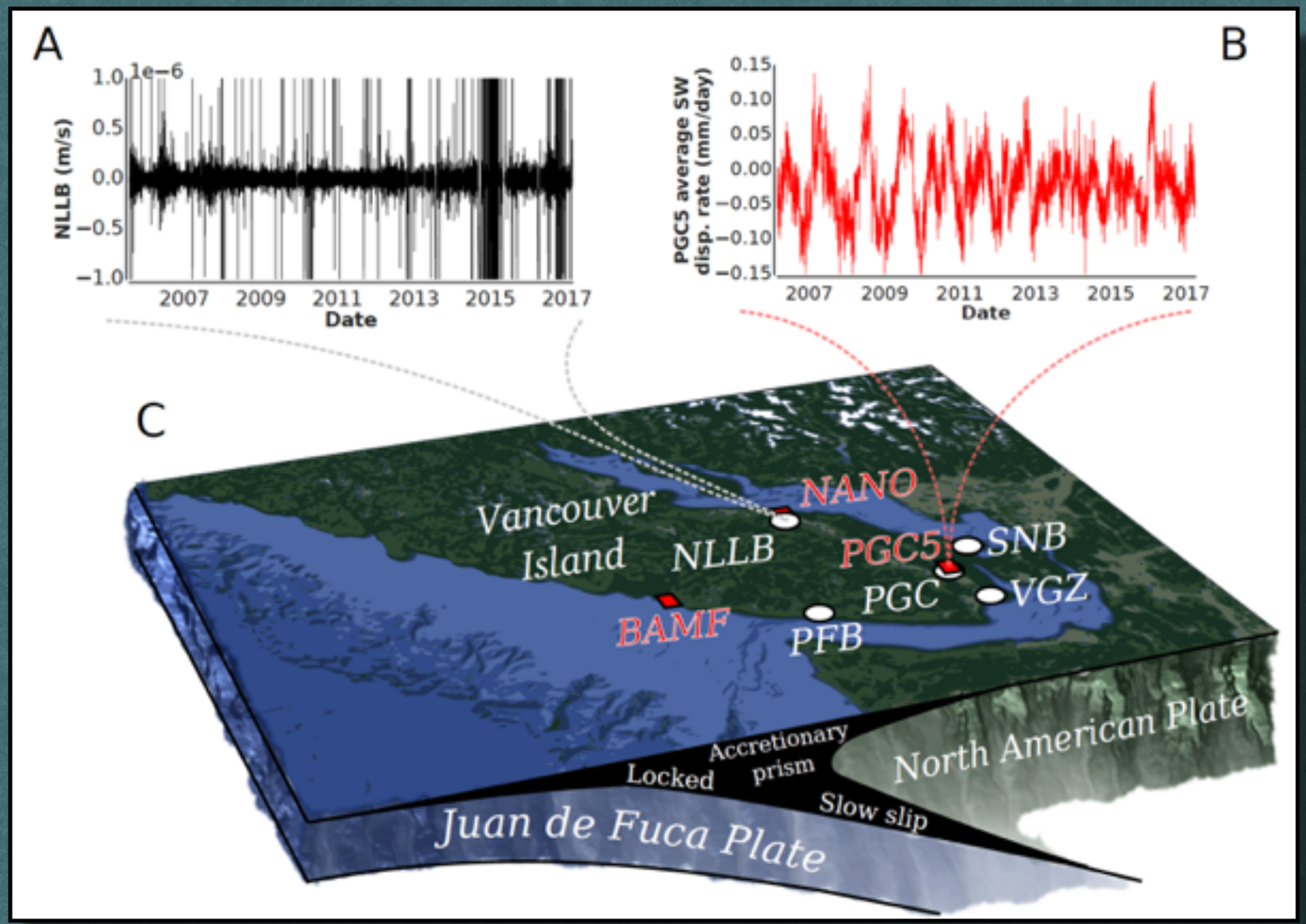
UNKNOWN SIGNAL REVEALED BY ML



The tremor tells us the gouge is mobile and chattering, post failure. Very quickly after a quake, the material has re-arranged itself in preparation for the quake.

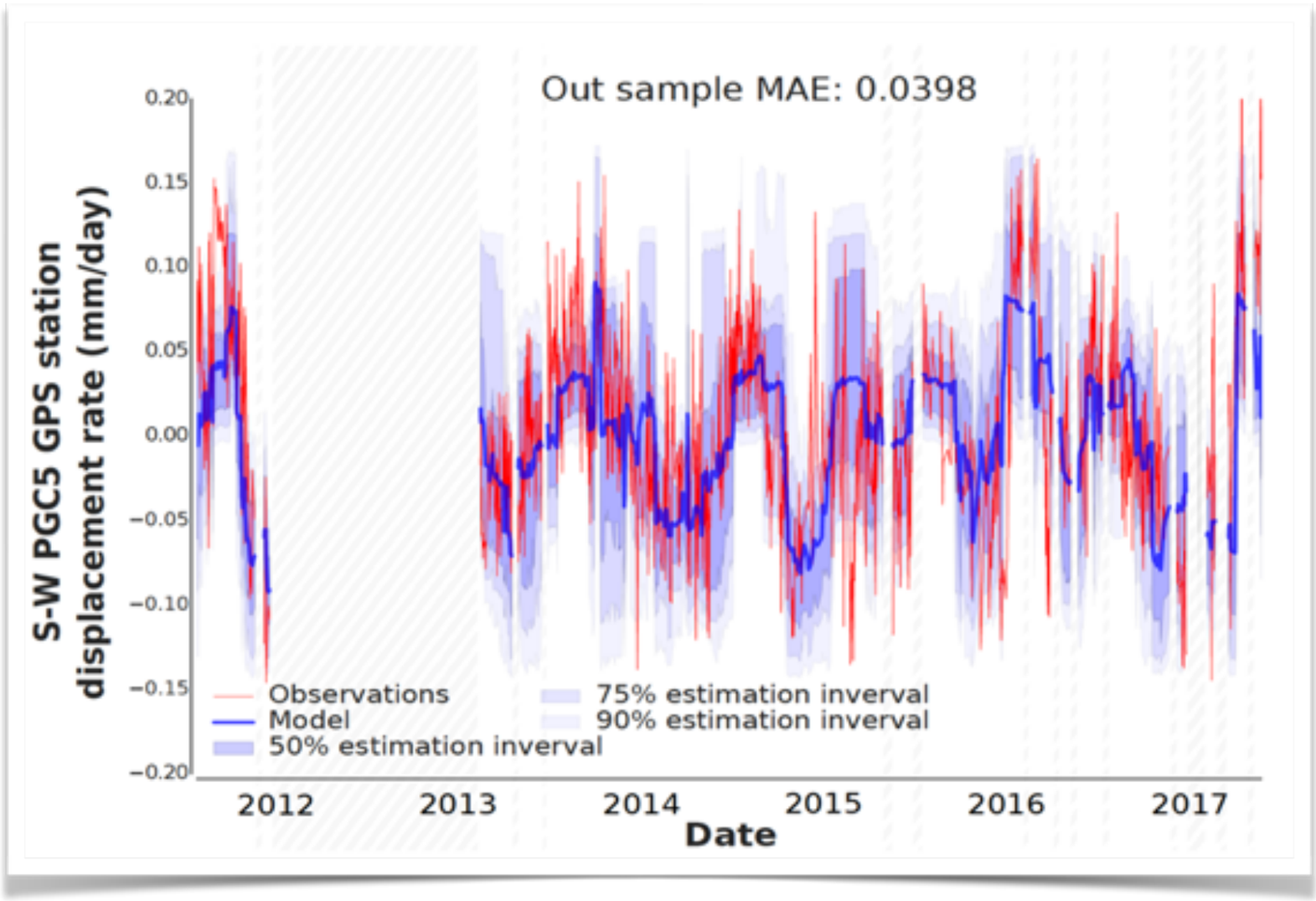
Rouet-LeDuc, Geophys. Res. Lett, 2017

SCALE TO EARTH? SLOW SLIP IN CASCADIA



Rouet-LeDuc, Hulbert & PAJ, Nature Geosciences, 2019

CASCADIA DISPLACEMENT RATE



Rouet-LeDuc, Hulbert & PAJ, Nature Geosciences, 2019