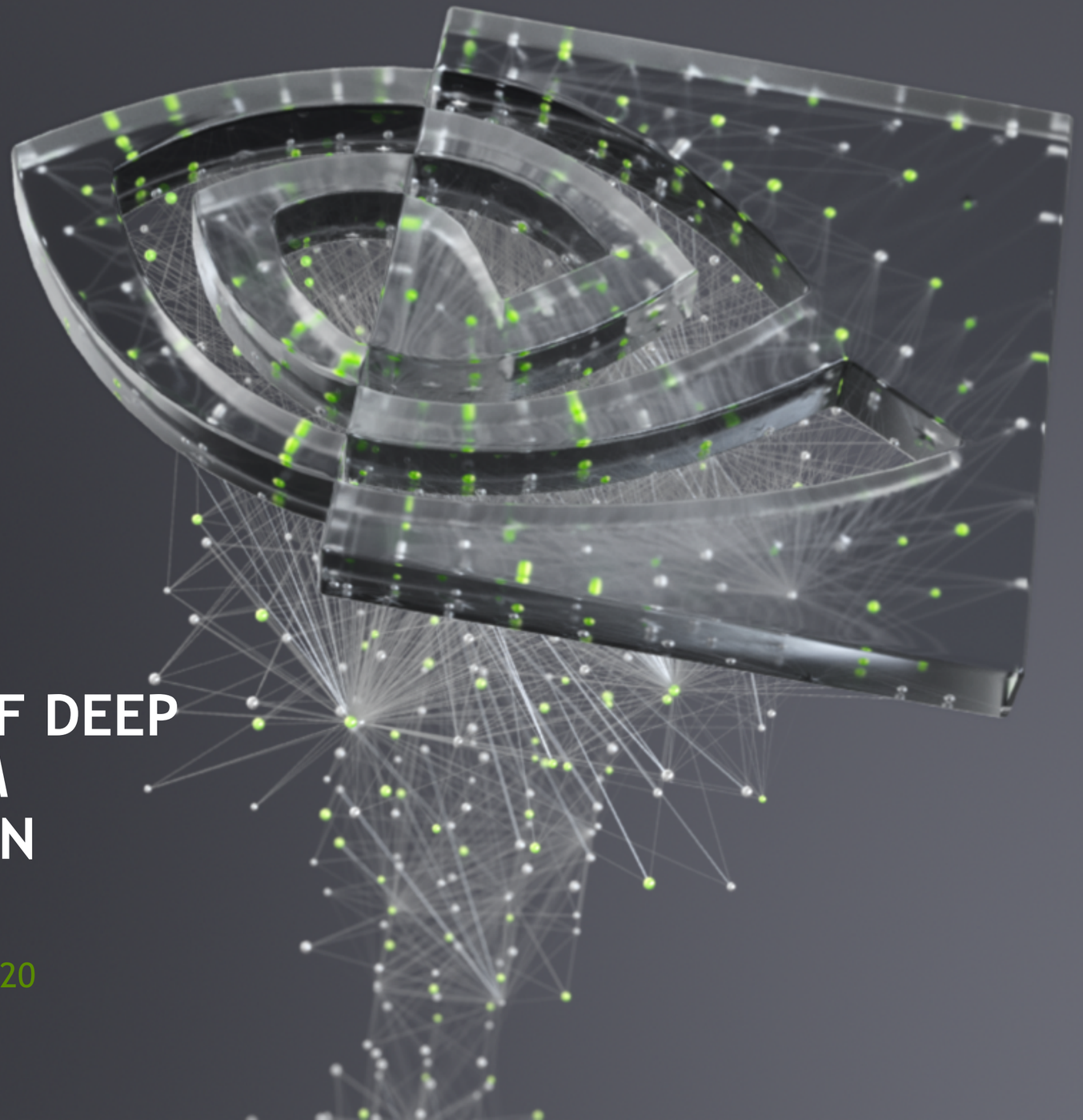


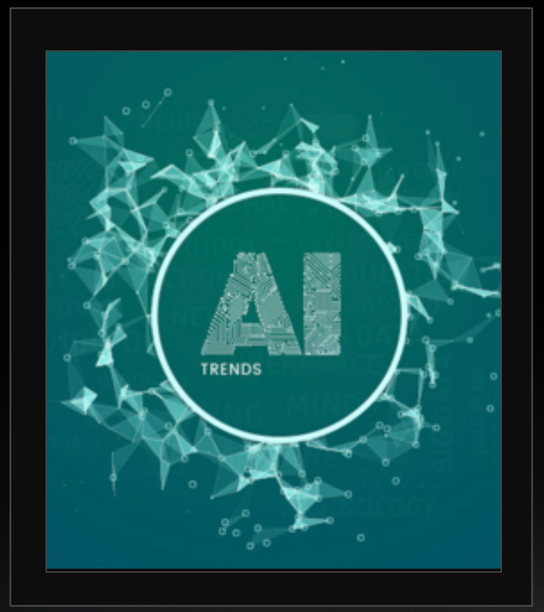
EXPLORING THE FRONTIERS OF DEEP LEARNING FOR EARTH SYSTEM OBSERVATION AND PREDICTION

Dr. David M. Hall, Senior Data Scientist, NVIDIA

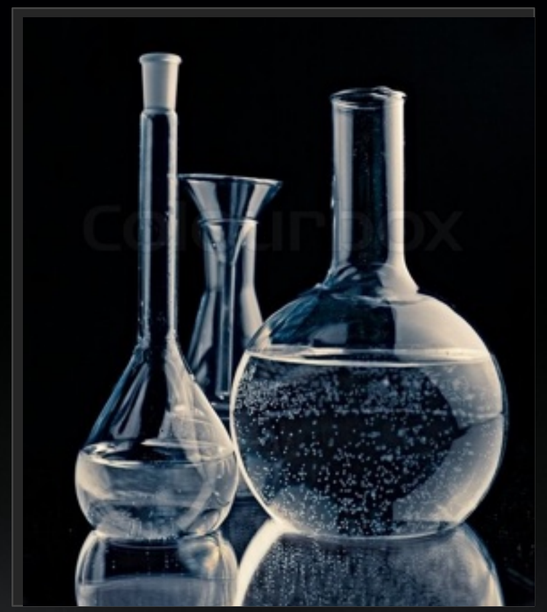
ECMWF-ESA Workshop on Machine Learning, Oct, 2020

THE FRONTIERS OF DEEP LEARNING

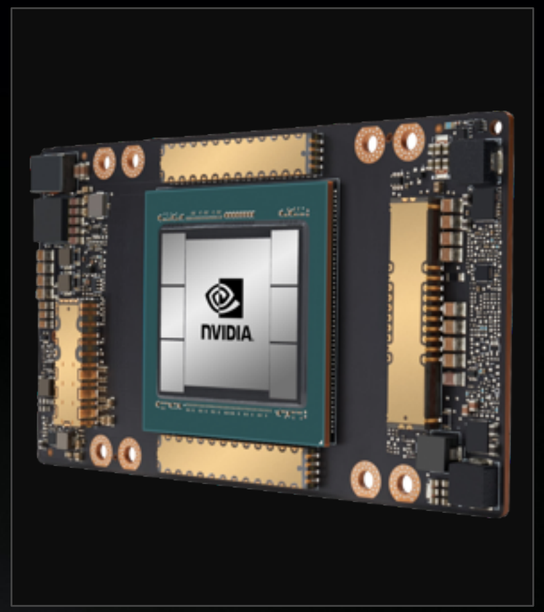
Applications



AI Trends



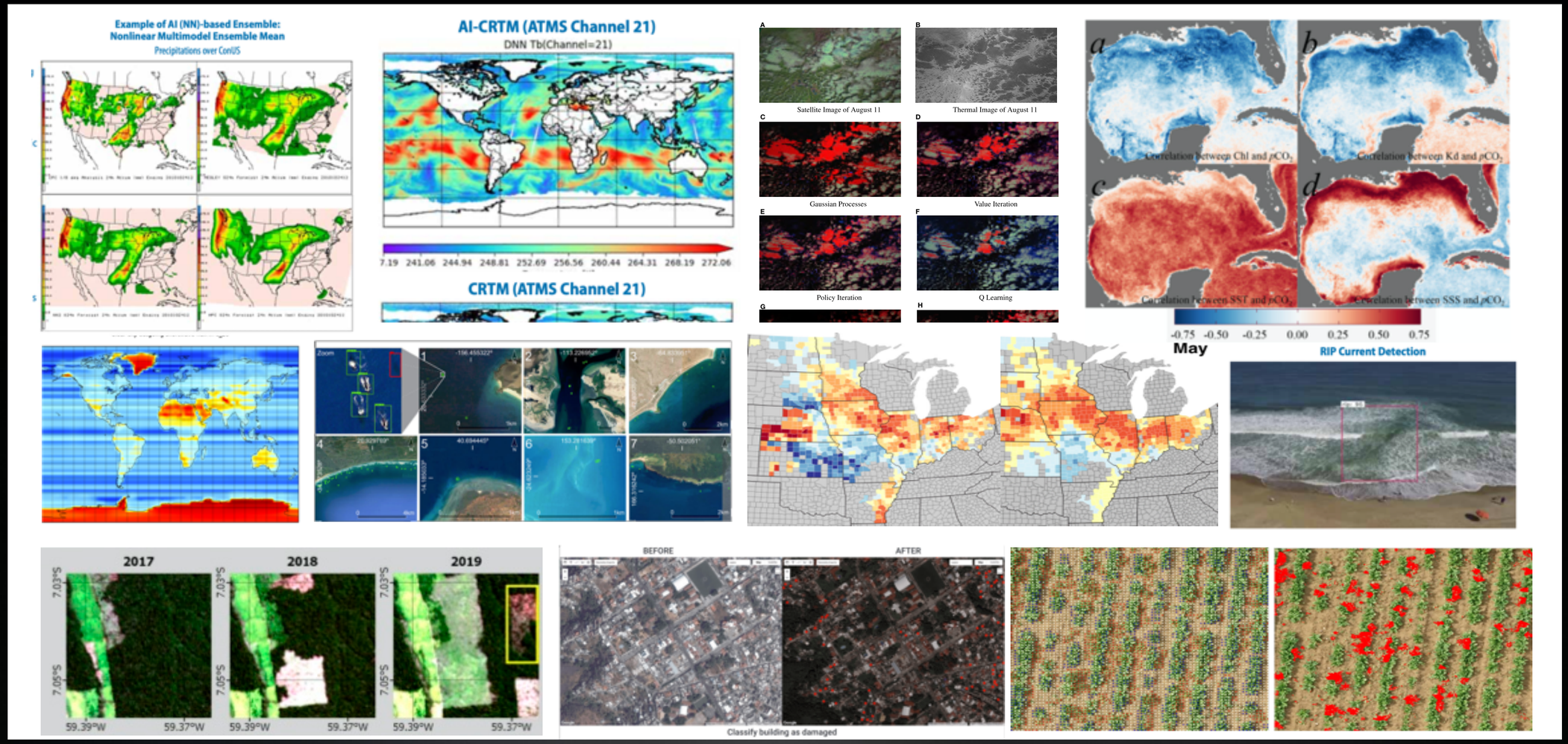
Scientific Challenges



Hardware and Tools

RAPID ADOPTION

Deep learning is being rapidly adopted by the Earth System Science community



NSF AI INSTITUTES

Seven new Artificial Intelligence institutes, including one for Weather and Climate

NSF AI Institute for Research on Trustworthy AI in Weather, Climate and Coastal Oceanography

NSF announcement

NOAA CENTER FOR AI

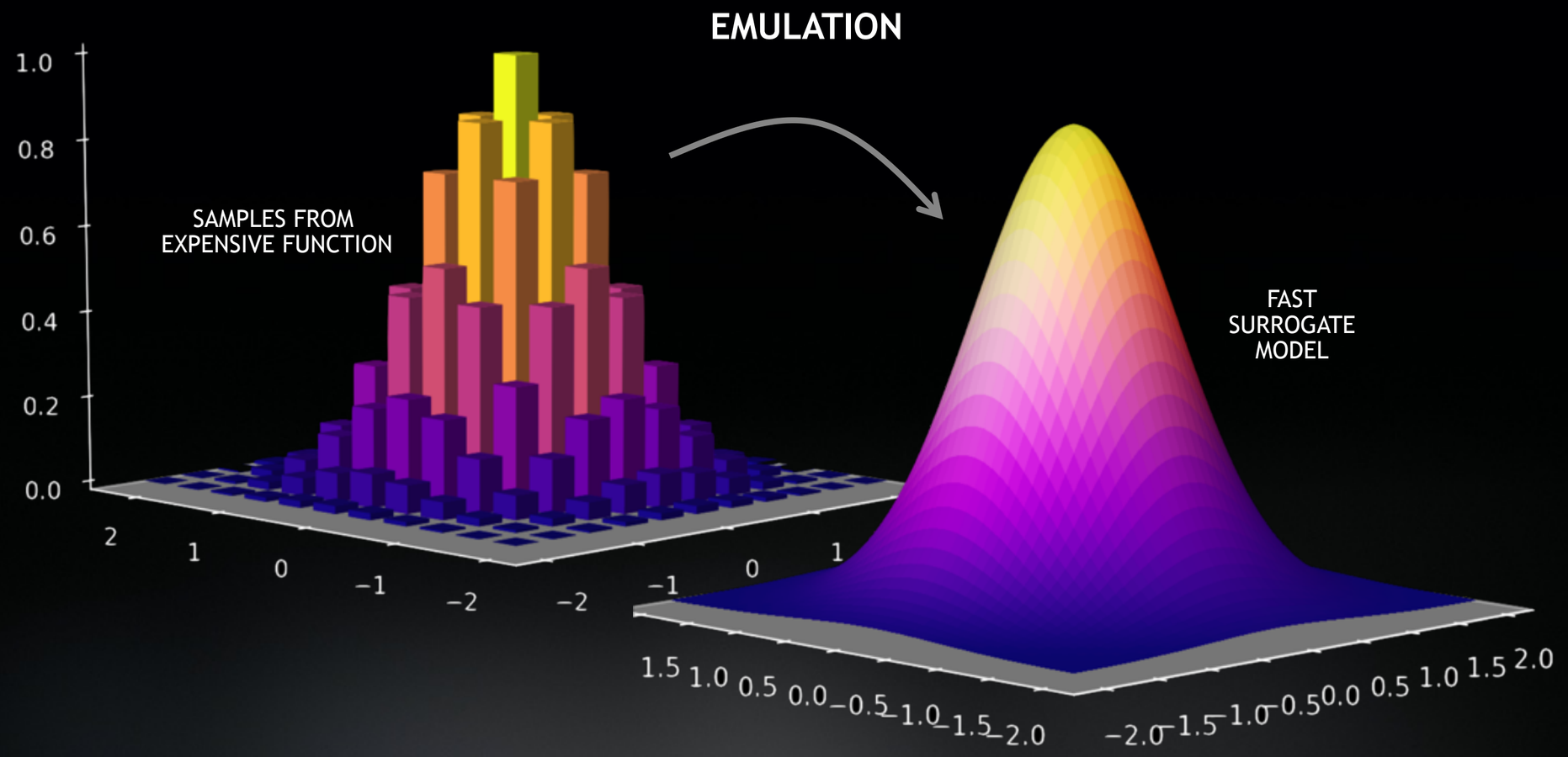
Official strategy and dedicated center focusing on AI

https://nrc.noaa.gov/LinkClick.aspx?fileticket=0I2p2-Gu3rA%3D&tabid=91&portalid=0

APPLICATIONS IN EARTH SYSTEM SCIENCE

ACCELERATED PHYSICS

Using surrogate models to speed up existing code



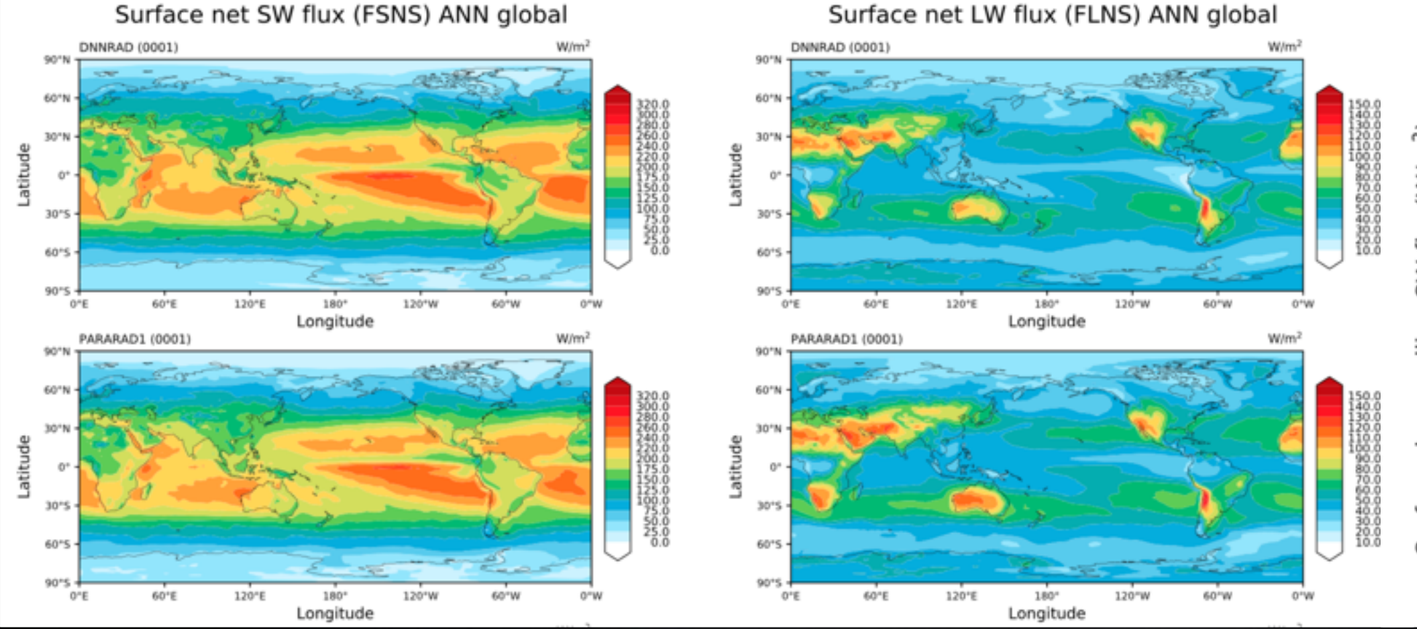
ACCELERATED PHYSICS PARAMETERIZATIONS

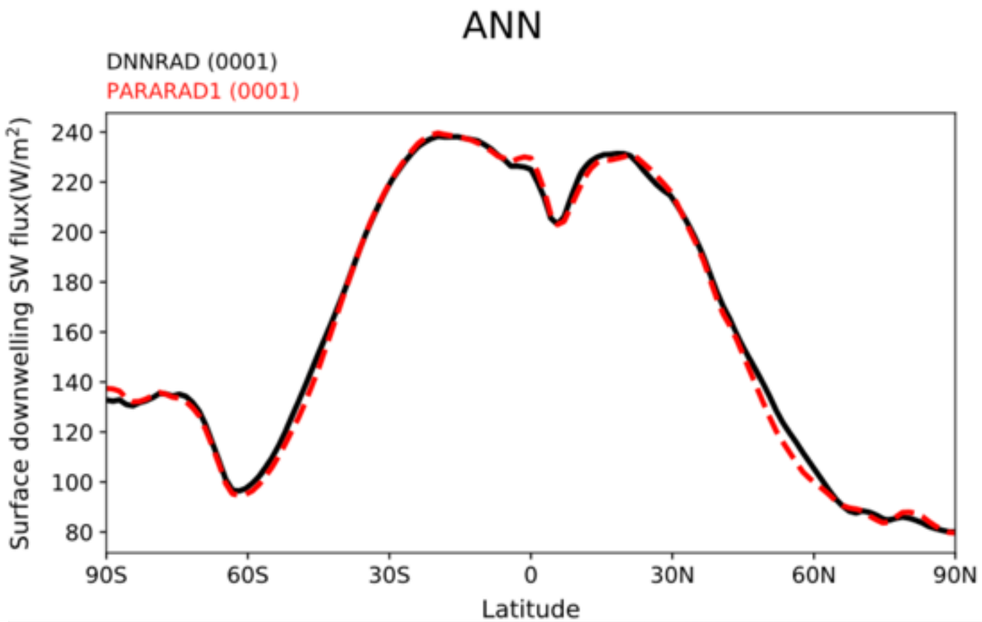
Emulation of E3SM Super-parameterized SW and LW Radiation

Using Deep Neural Networks as Cost-Effective Surrogate Models for Super-Parameterized E3SM Radiative Transfer

Anikesh Pal[1], Salil Mahajan[2], and Matthew R. Norman[1]

[1]National Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, TN, USA, [2]Computational Earth SciencesOak Ridge National Laboratory, Oak Ridge, TN, USA





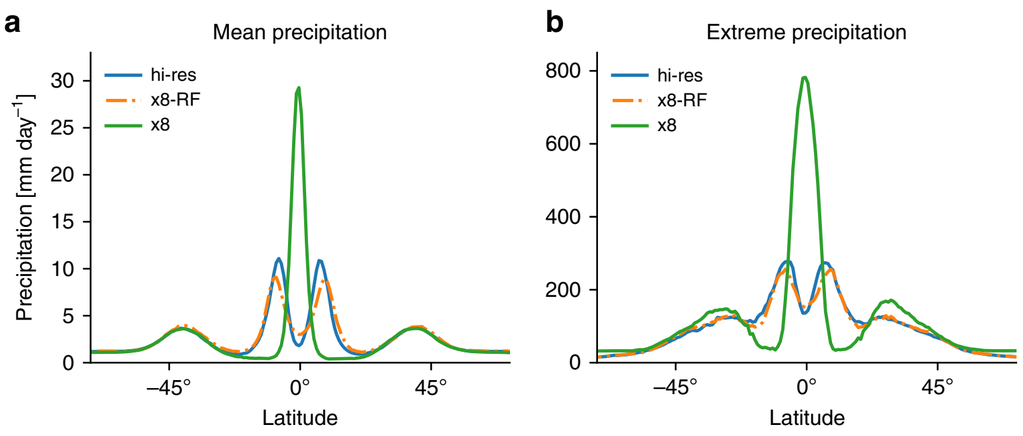
8-10x Speedup in SW and LW Radiative Transfer Calculations

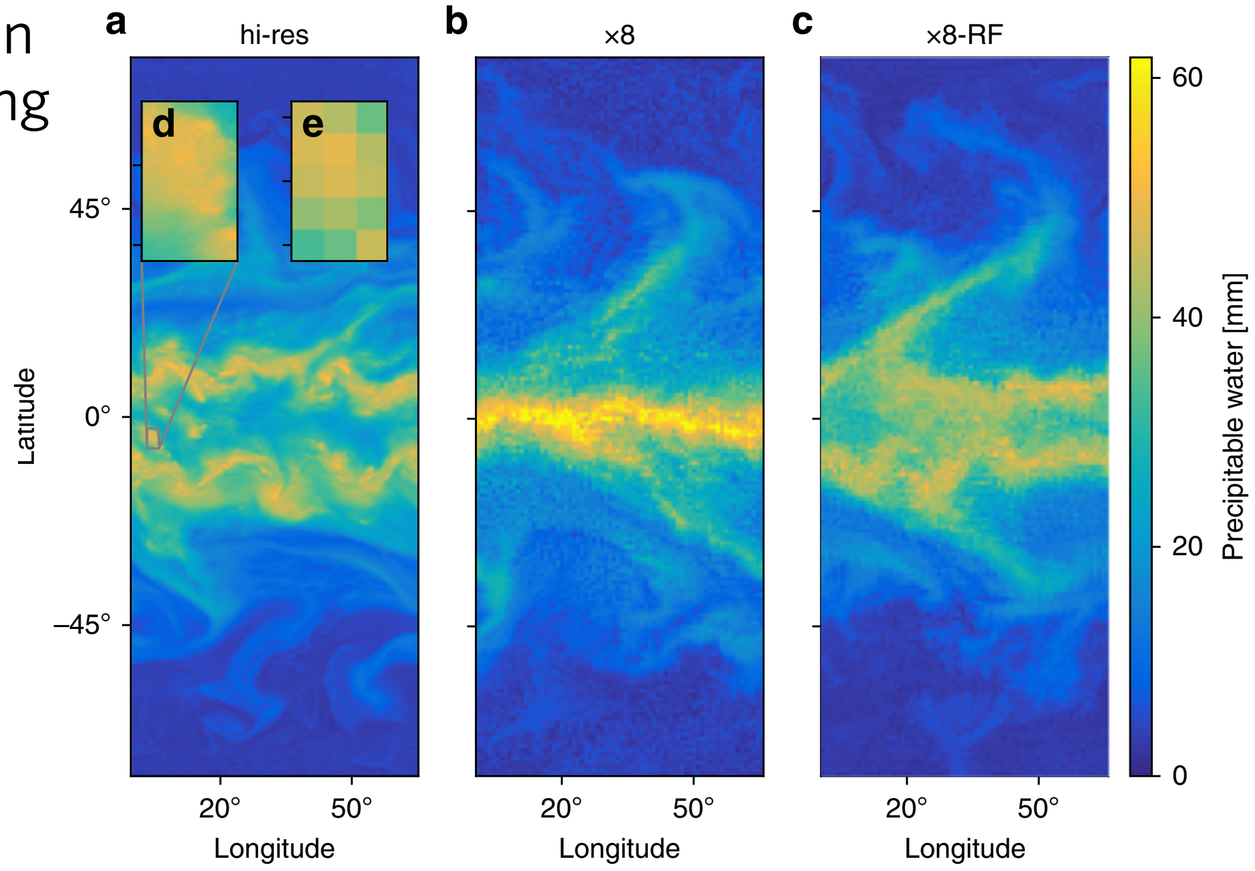
IMPROVED PHYSICS PARAMETERIZATIONS

Building more accurate physics parameterizations

Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions

Janni Yuval[1] & Paul A. O'Gorman[1]

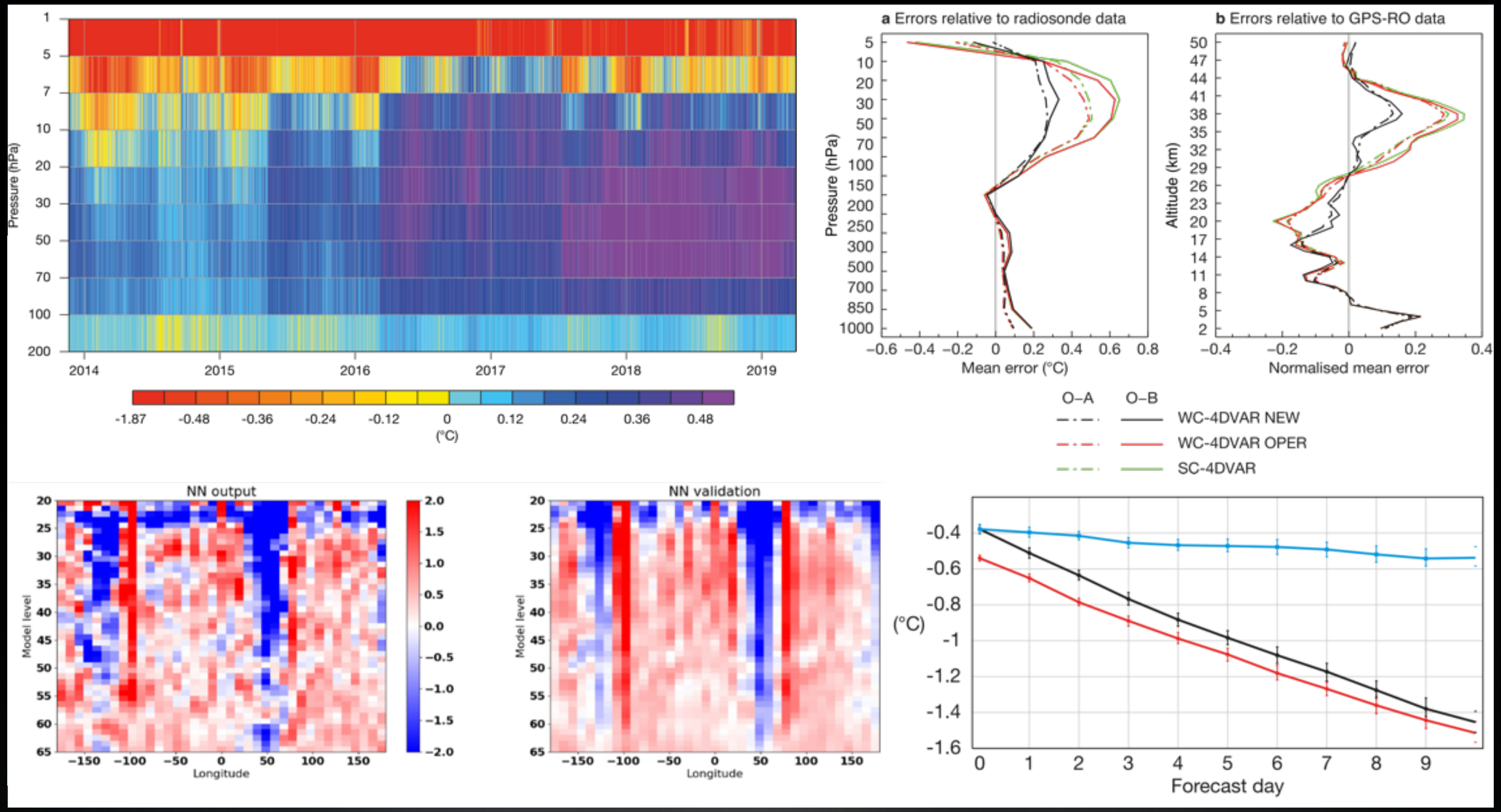




https://www.nature.com/articles/s41467-020-17142-3

BIAS CORRECTION

Work with ECMWF to remove IFS Model bias

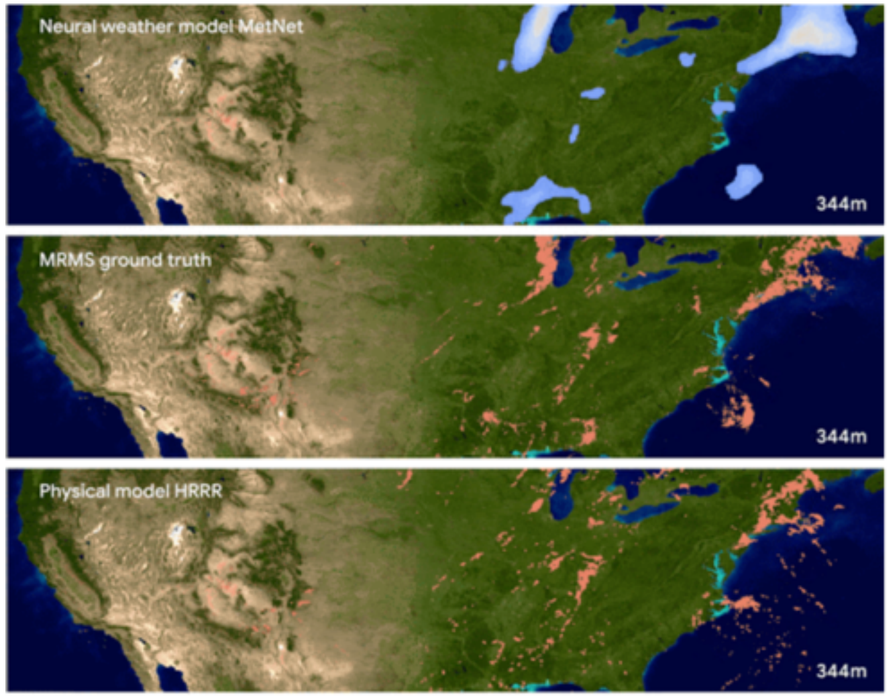


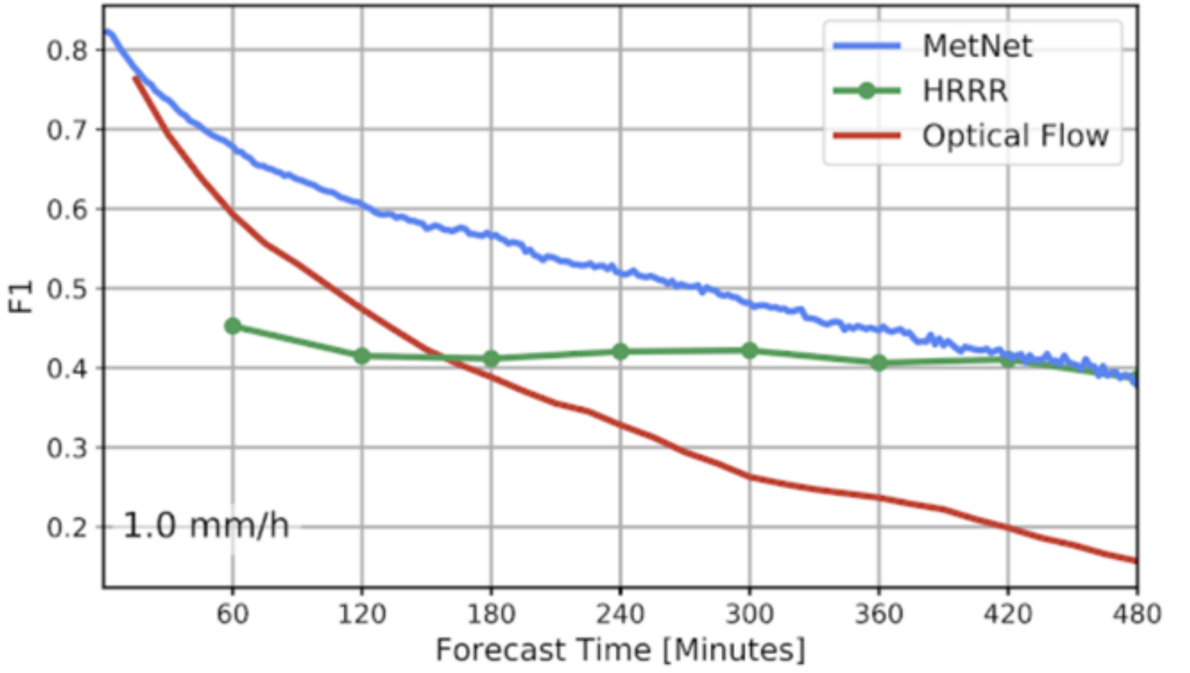
http://dx.doi.org/10.21957/pl881qs63d

NOWCASTING

MetNet: Can beat physical models up to 8 hours

MetNet: A Neural Weather Model for Precipitation Forecasting GOOGLE RESEARCH





Performance evaluated in terms of F1-score at 1.0 mm/h precipitation rate (higher is better). The neural weather model (MetNet) outperforms the physics-based model (HRRR) currently operational in the US for timescales up to 8 hours ahead.

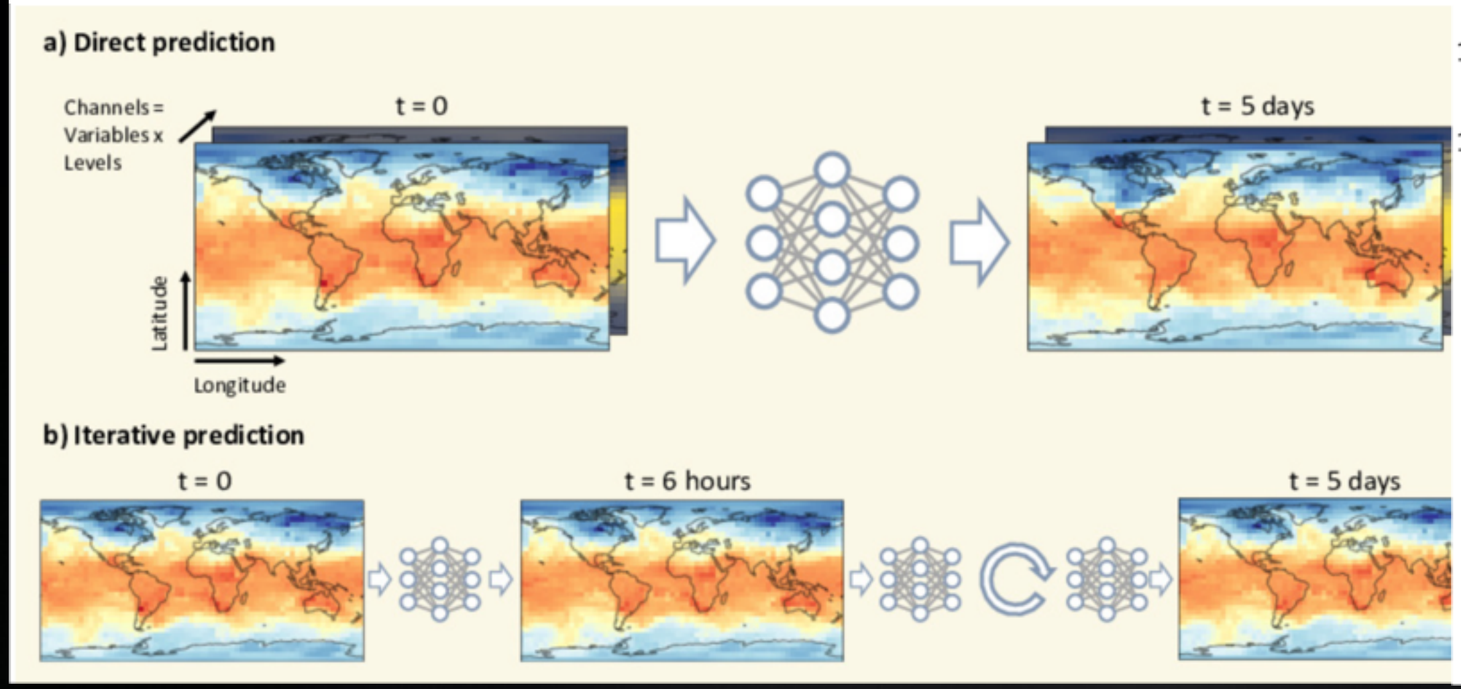
https://ai.googleblog.com/2020/03/a-neural-weather-model-for-eight-hour.html

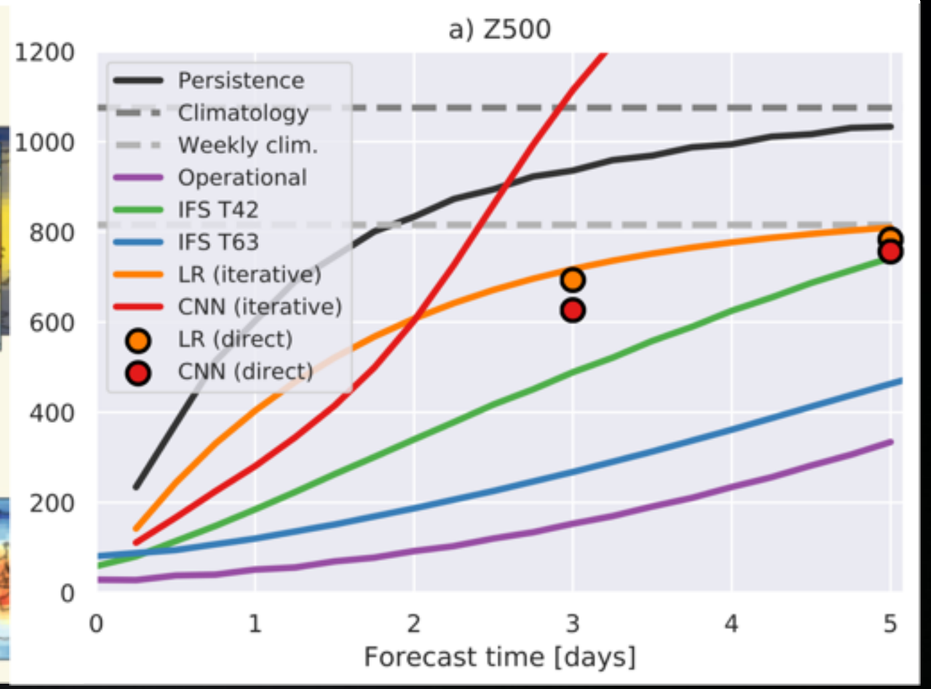
MEDIUM-RANGE WEATHER FORECASTING

Weather Bench: A Standardized Benchmark for Data Driven Forecasts

WeatherBench: A benchmark dataset for data-driven weather forecasting

Stephan Rasp, Peter D. Dueben, Sebastian Scher, Jonathan A. Weyn, Soukayna Mouatadid, Nils Thuerey





https://arxiv.org/abs/2002.00469

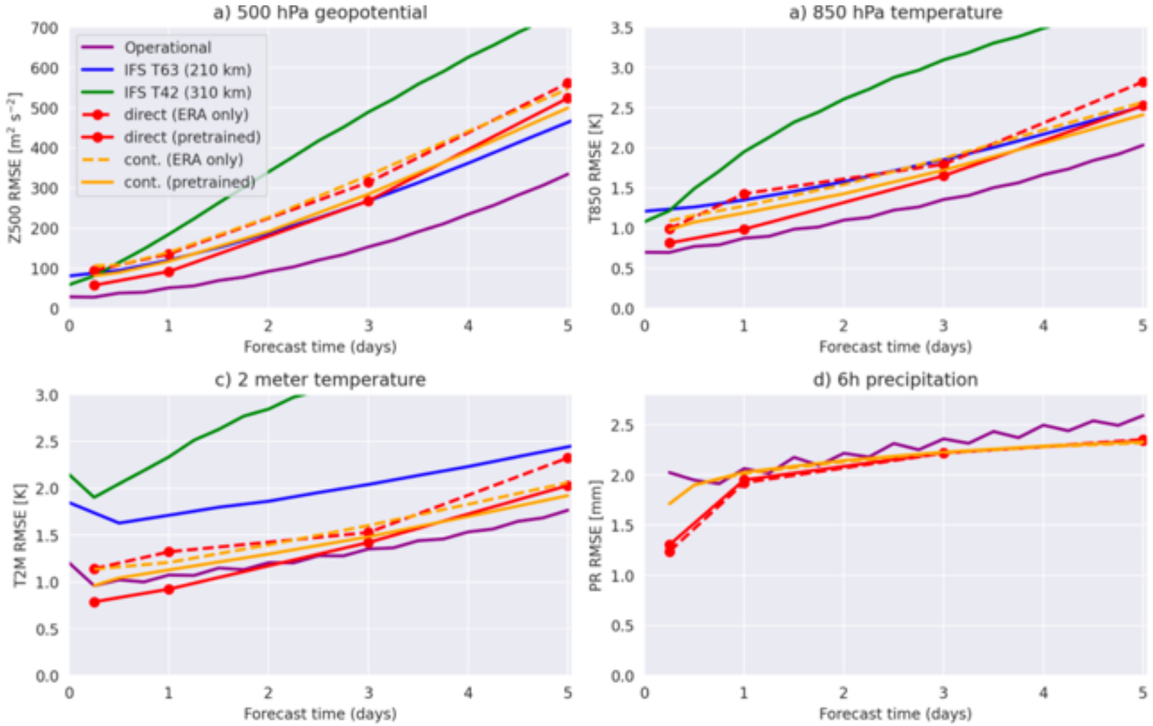
MEDIUM-RANGE WEATHER FORECASTING

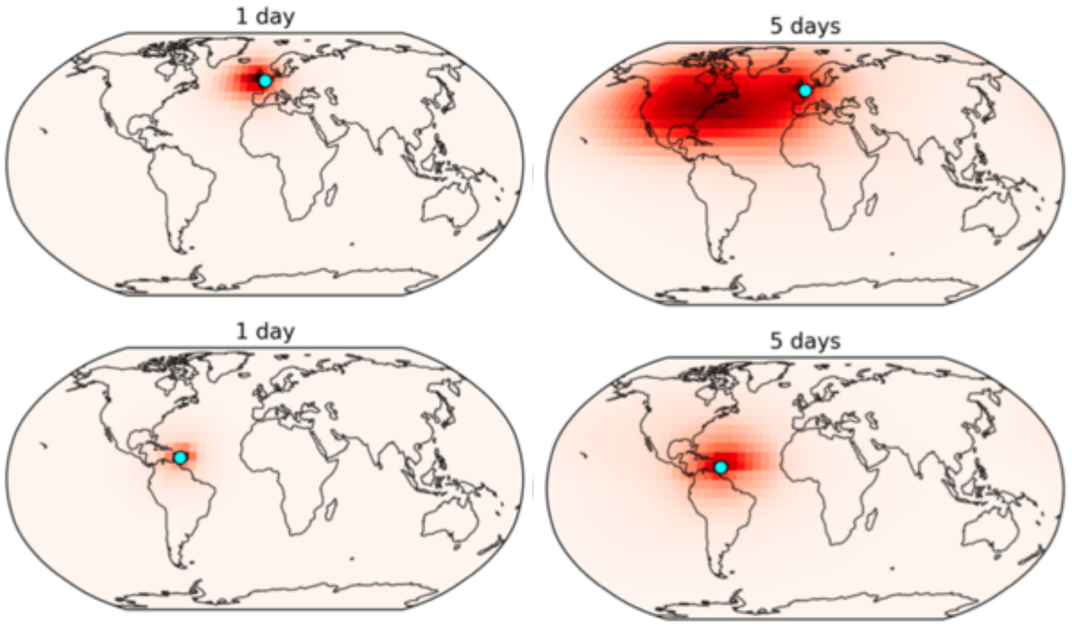
Data Driven models match accuracy of physical models on weather-bench

Purely data-driven medium-range weather forecasting achieves comparable skill to physical models at similar resolution

Stephan Rasp
Department of Informatics
Technical University of Munich
Munich, Germany
stephan.rasp@tum.de

Nils Thuerey
Department of Informatics
Technical University of Munich
Munich, Germany
nils.thuerey@tum.de



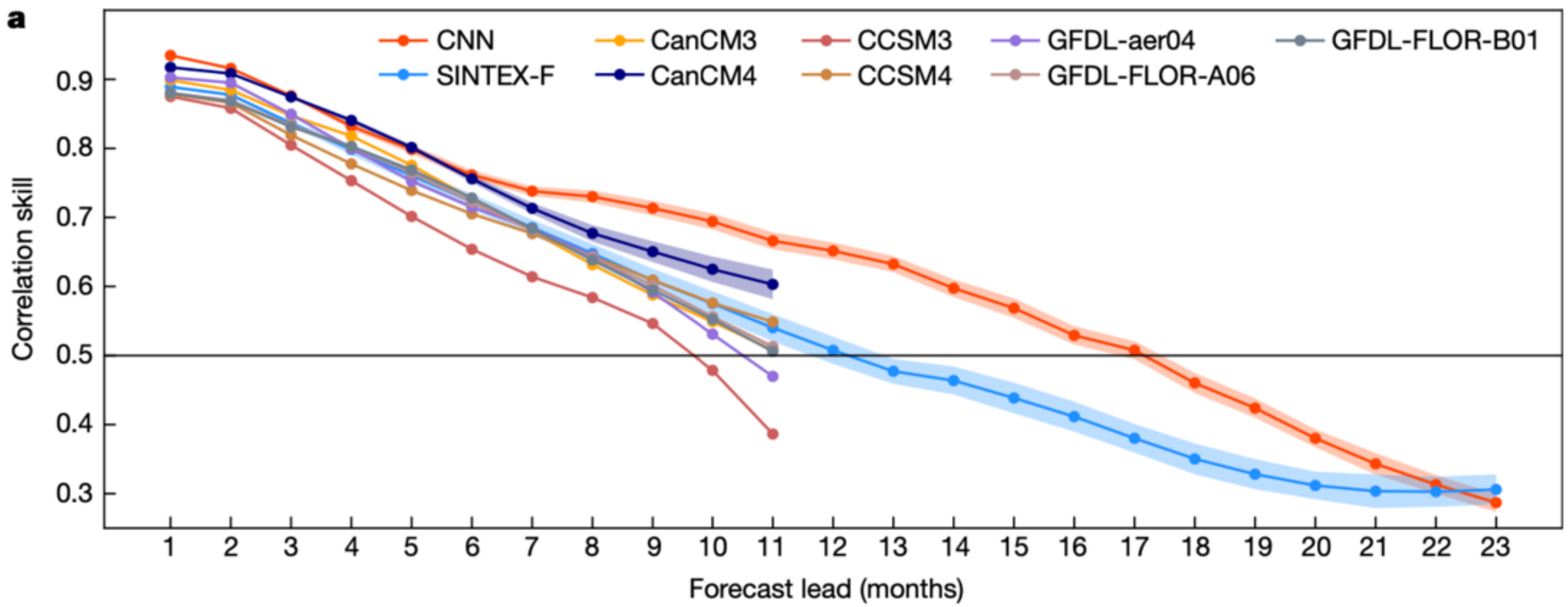


EL NIÑO PREDICTION

Deep learning model able to predict El Niño up to 18 months in advance

Deep learning for multi-year ENSO forecasts

Yoo-Geun Ham[1]*, Jeong-Hwan Kim[1] & Jing-Jia Luo[2,3]



https://www.nature.com/articles/s41586-019-1559-7.pdf

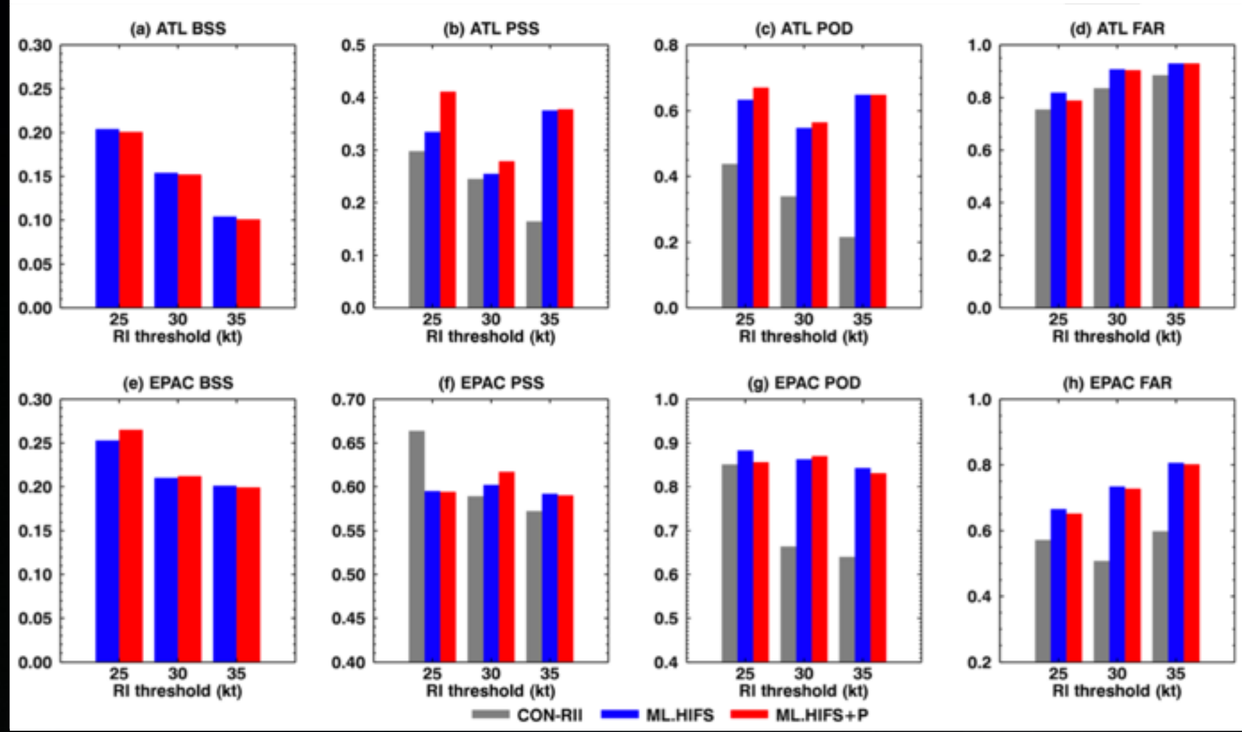
RAPID INTENSIFICATION

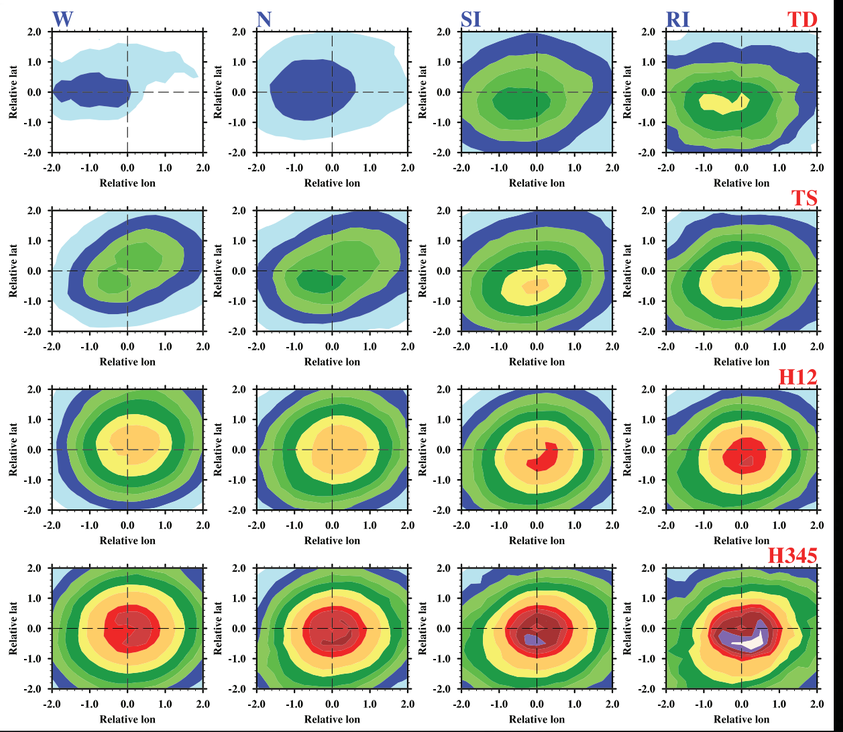
Machine Learning improves detection by 40-200 % over operational consensus

Applying Satellite Observations of Tropical Cyclone Internal Structures to Rapid Intensification Forecast With Machine Learning

Hui Su[1], Longtao Wu[1], Jonathan H. Jiang[1], Raksha Pai[2], Alex Liu[3], Albert J. Zhai[4], Peyman Tavallali[1], and Mark DeMaria[5]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, [2]IBM Global Business Services, Armonk, NY, USA, [3]RMDS Lab, Pasadena, CA, USA, [4]Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA, [5]National Hurricane Center, Miami, FL, USA





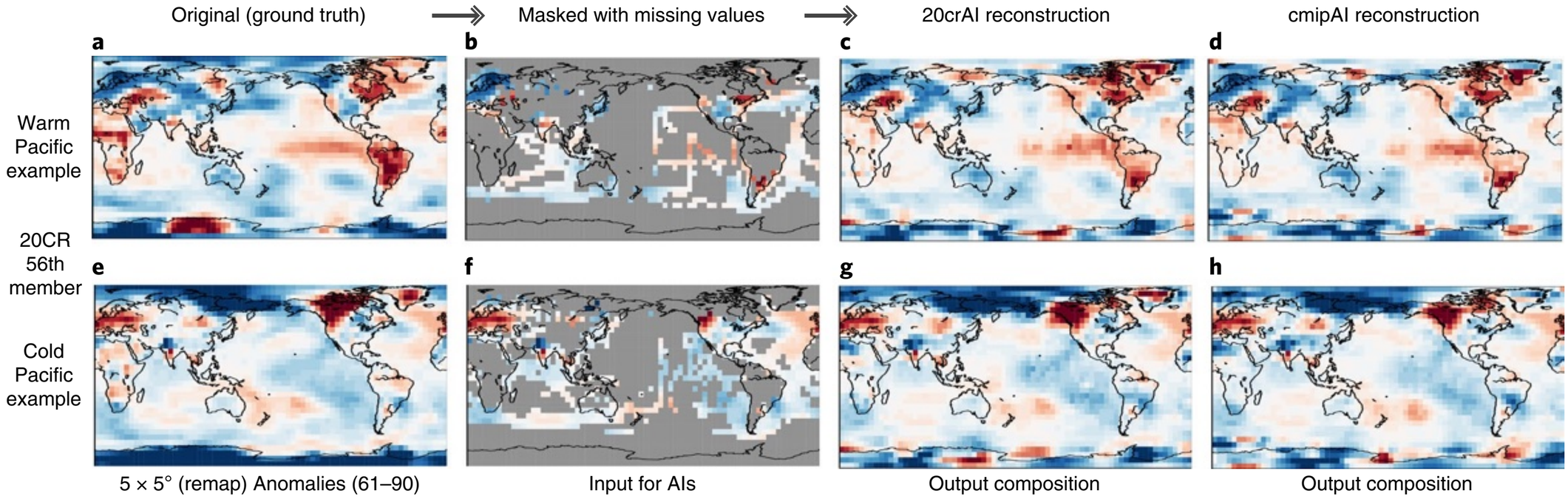
https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2020GL089102

FILLING-IN MISSING CLIMATE OBSERVATIONS

DL transfer learning + inpainting beats Kriging and PCA

Artificial intelligence reconstructs missing climate information

Christopher Kadow[1,2] ✉, David Matthew Hall[3] and Uwe Ulbrich[2]



SCIENTIFIC CHALLENGES

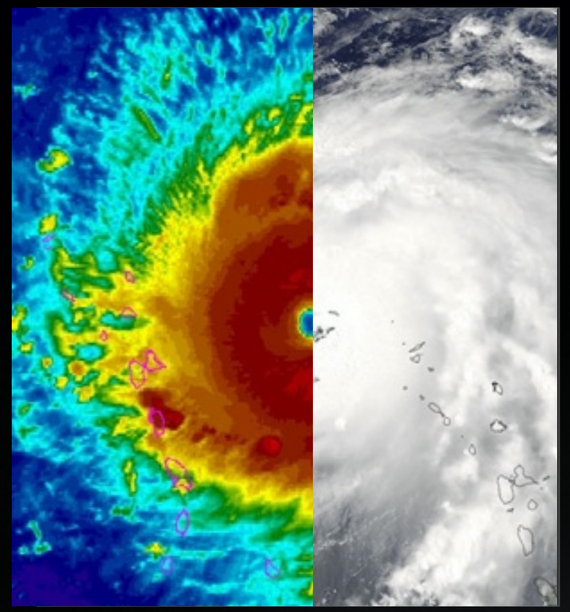
SCIENTIFIC CHALLENGES

Problems that arise when applying AI to science

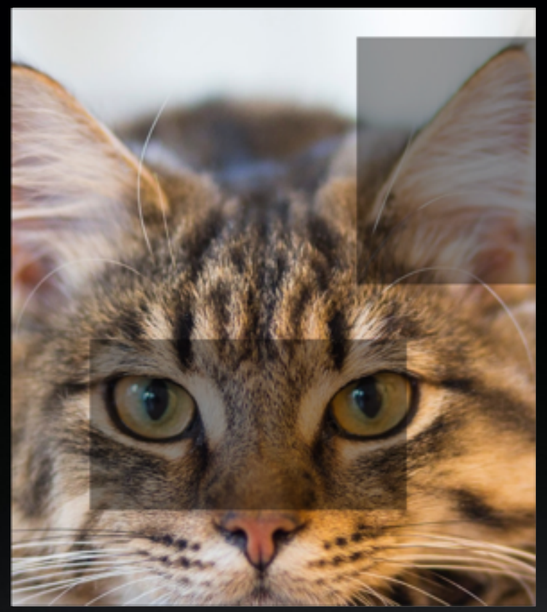
- Data Labelling
- Limited Data
- Enforcing Physical Constraints
- Uncertainty Quantification
- Trustworthiness and Interpretability
- HPC – AI Coupling
- Loss of Dynamic Range
- Data Movement
- Numerical Stability
- Generalization

DATA LABELLING

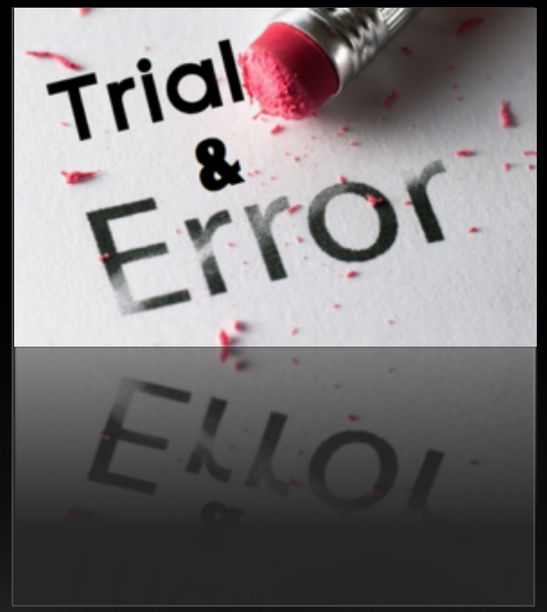
How can we get enough labelled data?



Data Fusion
Using one data source
as the label for another



Self-Supervised Learning
Predicting input B from input A



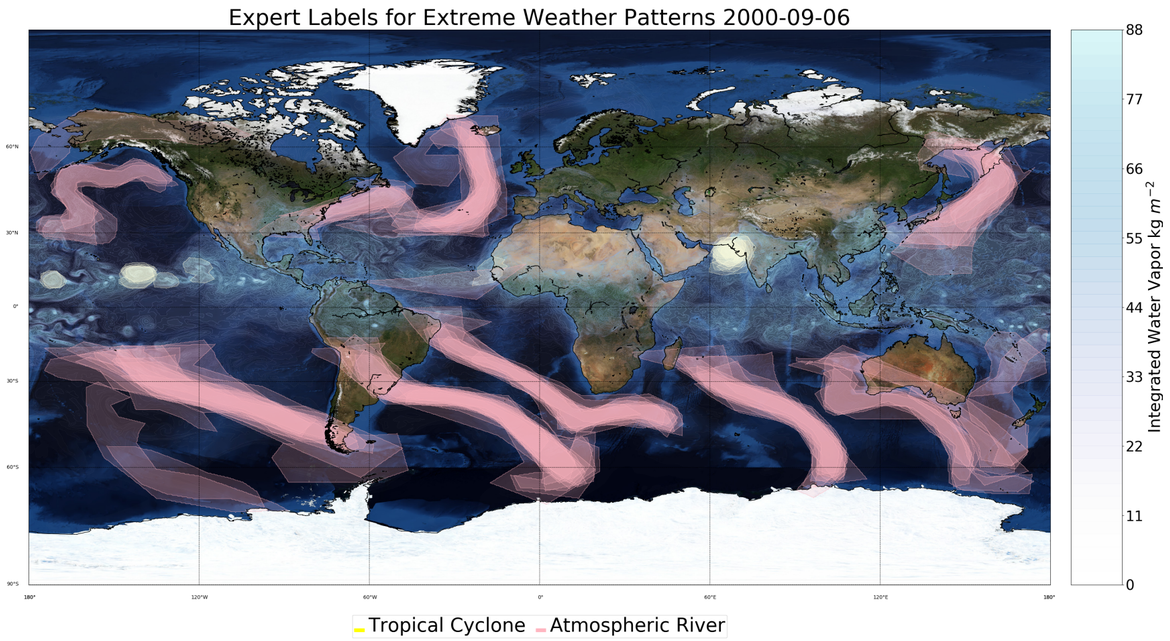
Reinforcement Learning
Obtaining labels directly from the
environment or simulation

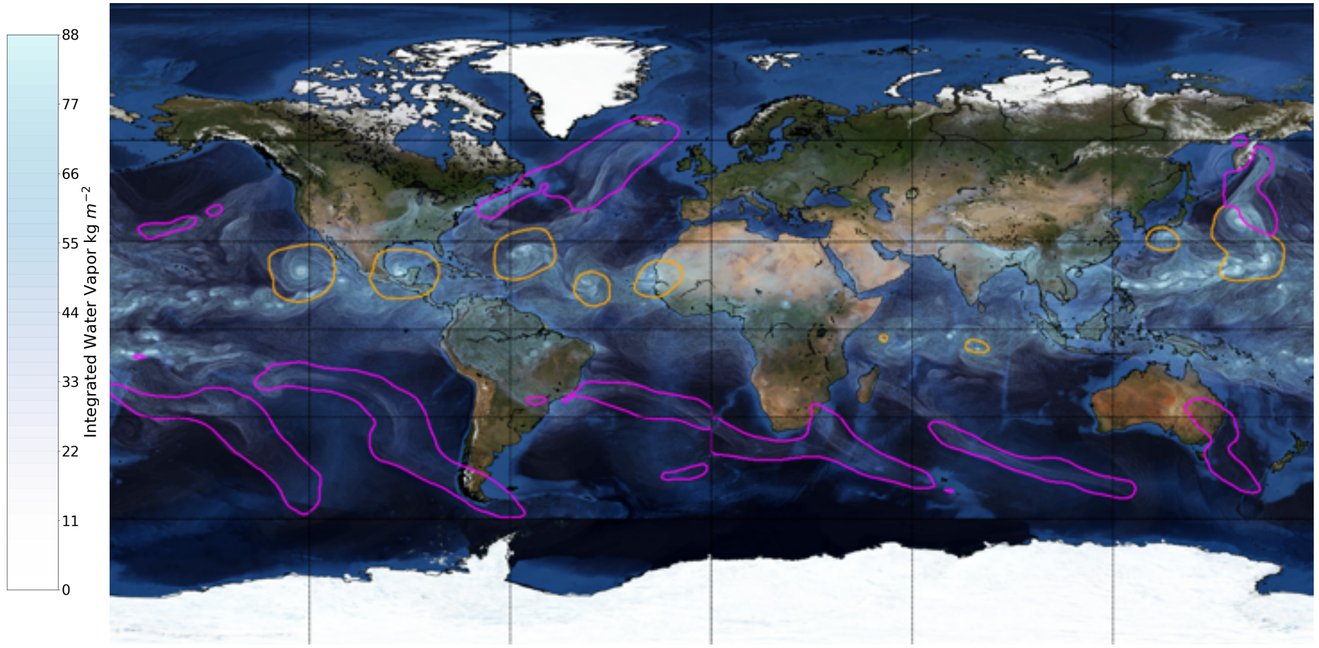
Active Learning
Using human machine iteration to
make labelling easier

DATA LABELLING

Online Extreme-Weather Labelling Tool

ClimateNet: an expert-labelled open dataset and Deep Learning architecture for enabling high-precision analyses of extreme weather

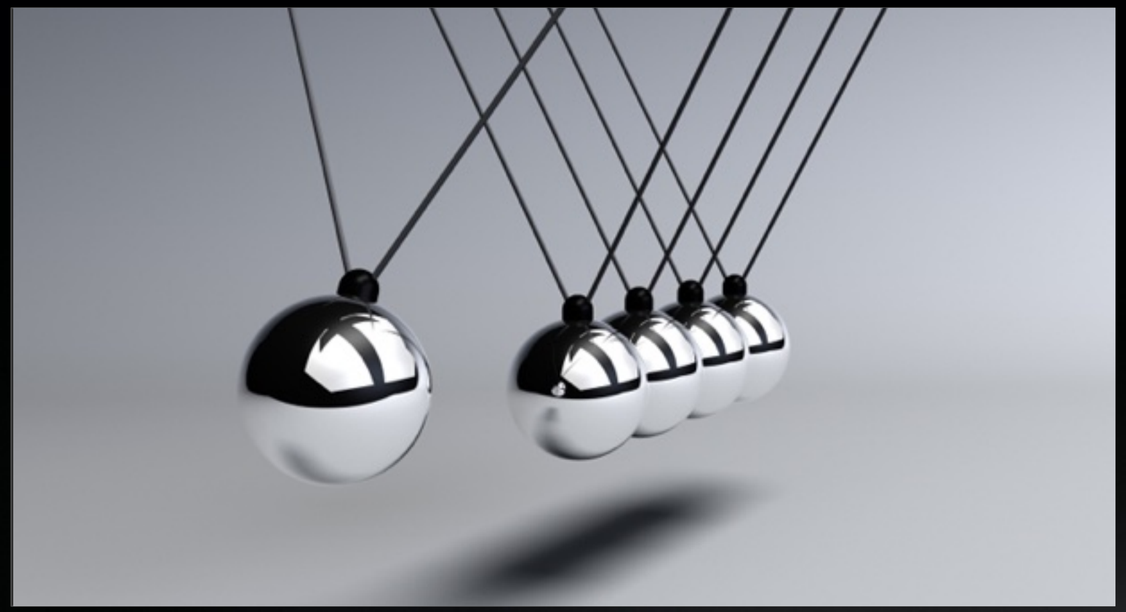




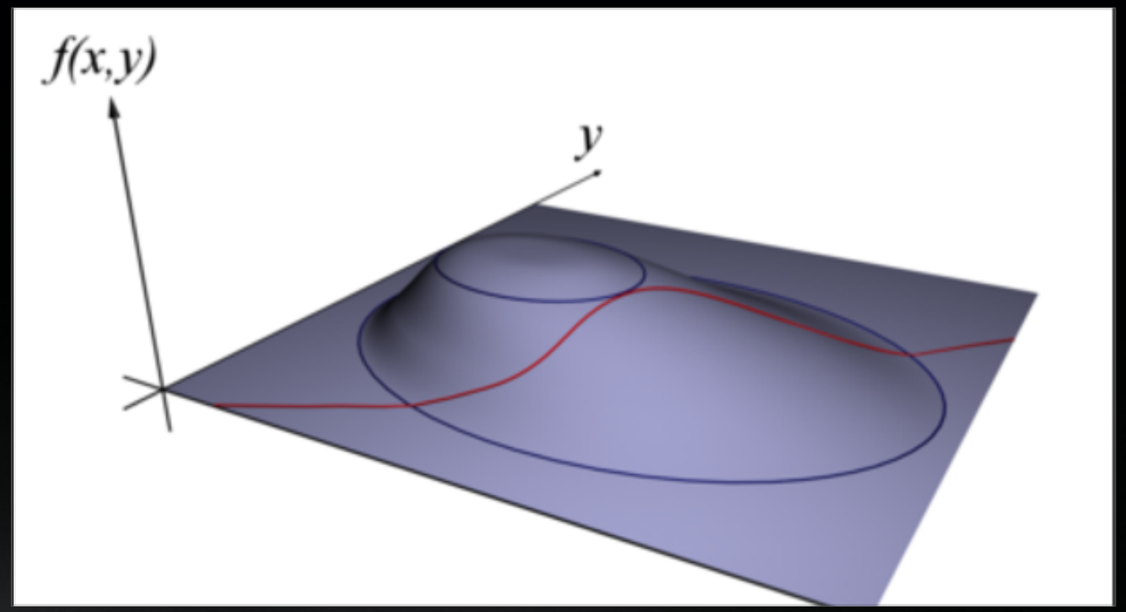
https://gmd.copernicus.org/preprints/gmd-2020-72/gmd-2020-72.pdf

ENFORCING PHYSICAL CONSTRAINTS

Is the solution physically correct?



Conservation of Mass, Momentum, Energy, Incompressibility, Turbulent Energy Spectra, Translational Invariance



Loss Penalization, Hard Constraints, Projective Methods, Differentiable Programs

ENFORCING PHYSICAL CONSTRAINTS

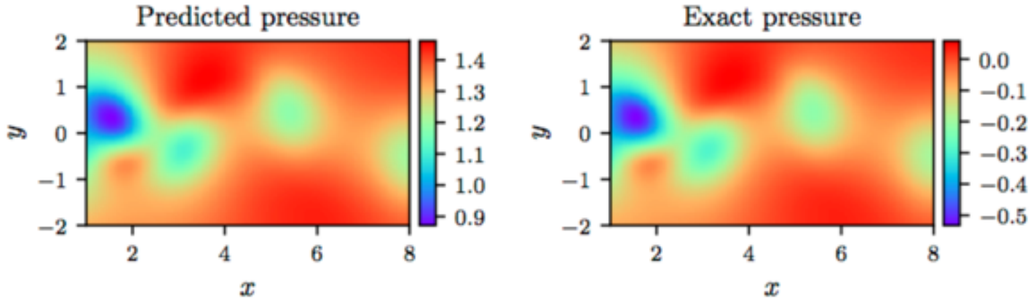
Physics Informed Neural Nets

Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations

M. Raissi [a], P. Perdikaris [b,*], G.E. Karniadakis [a]

[a] Division of Applied Mathematics, Brown University, Providence, RI, 02912, USA
[b] Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, PA, 19104, USA

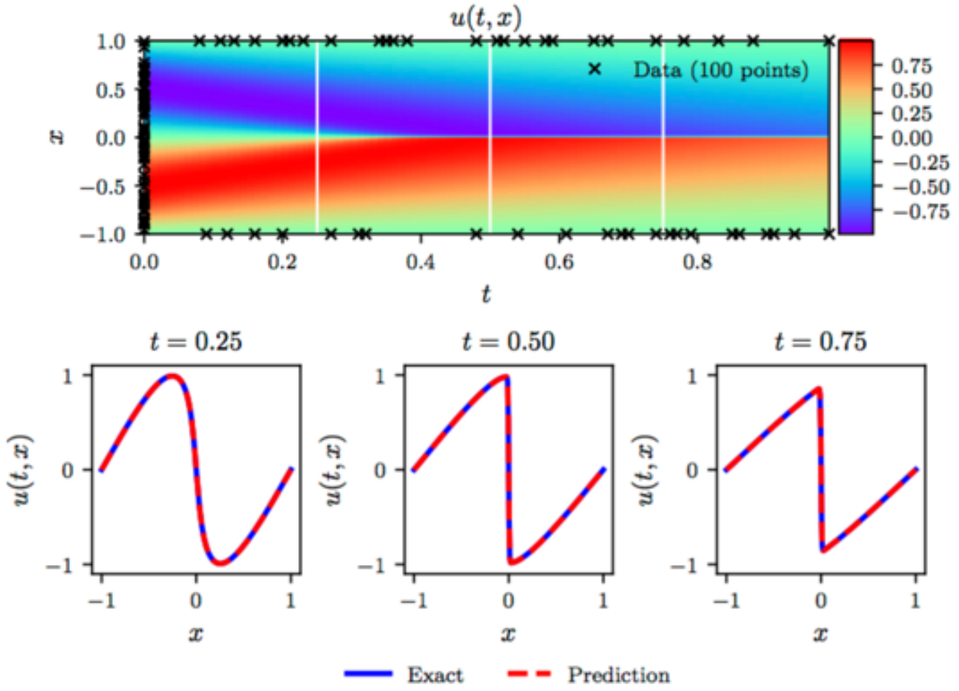
Data-driven Discovery of Nonlinear Partial Differential Equations



Correct PDE	$u_t + (uu_x + vu_y) = -p_x + 0.01(u_{xx} + u_{yy})$ $v_t + (uv_x + vv_y) = -p_y + 0.01(v_{xx} + v_{yy})$
Identified PDE (clean data)	$u_t + 0.999(uu_x + vu_y) = -p_x + 0.01047(u_{xx} + u_{yy})$ $v_t + 0.999(uv_x + vv_y) = -p_y + 0.01047(v_{xx} + v_{yy})$
Identified PDE (1% noise)	$u_t + 0.998(uu_x + vu_y) = -p_x + 0.01057(u_{xx} + u_{yy})$ $v_t + 0.998(uv_x + vv_y) = -p_y + 0.01057(v_{xx} + v_{yy})$

Data-driven Solutions of Nonlinear Partial Differential Equations

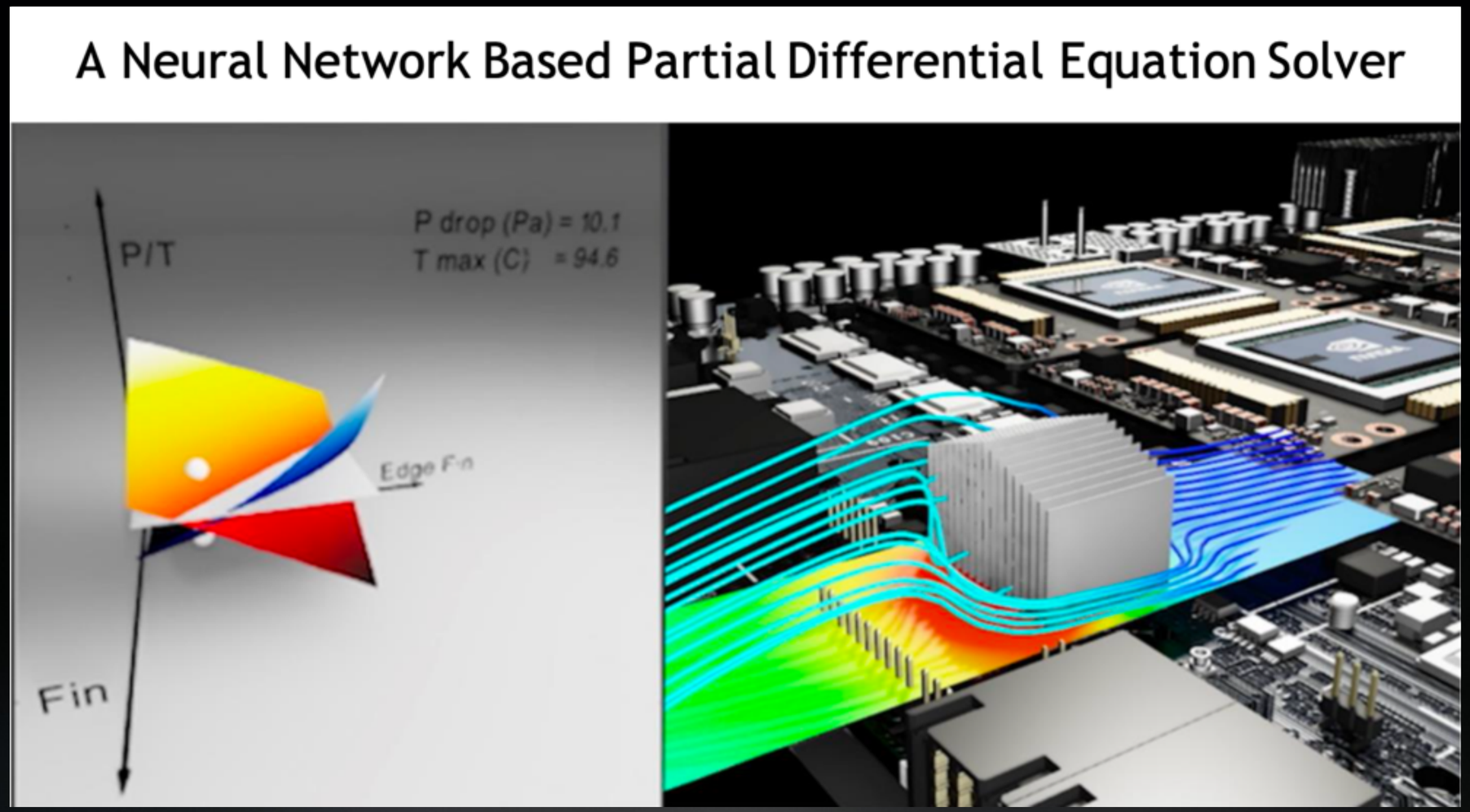
The following figure summarizes our results for the data-driven solution of the Burgers' equation.



https://maziarraissi.github.io/PINNs/

ENFORCING PHYSICAL CONSTRAINTS

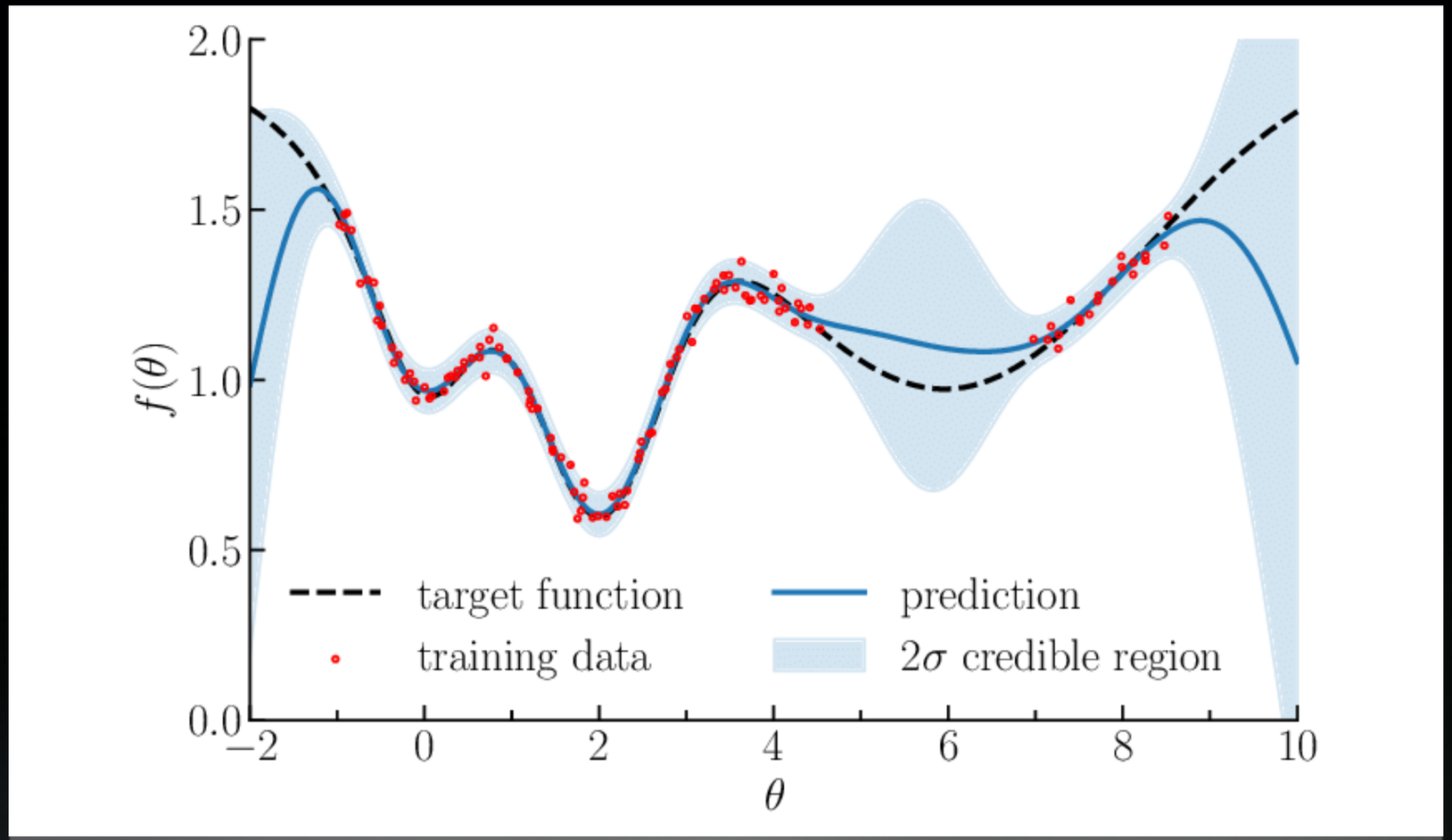
NVIDIA SimNet



https://www.youtube.com/watch?v=Oq2Mpi5pF1w

UNCERTAINTY QUANTIFICATION

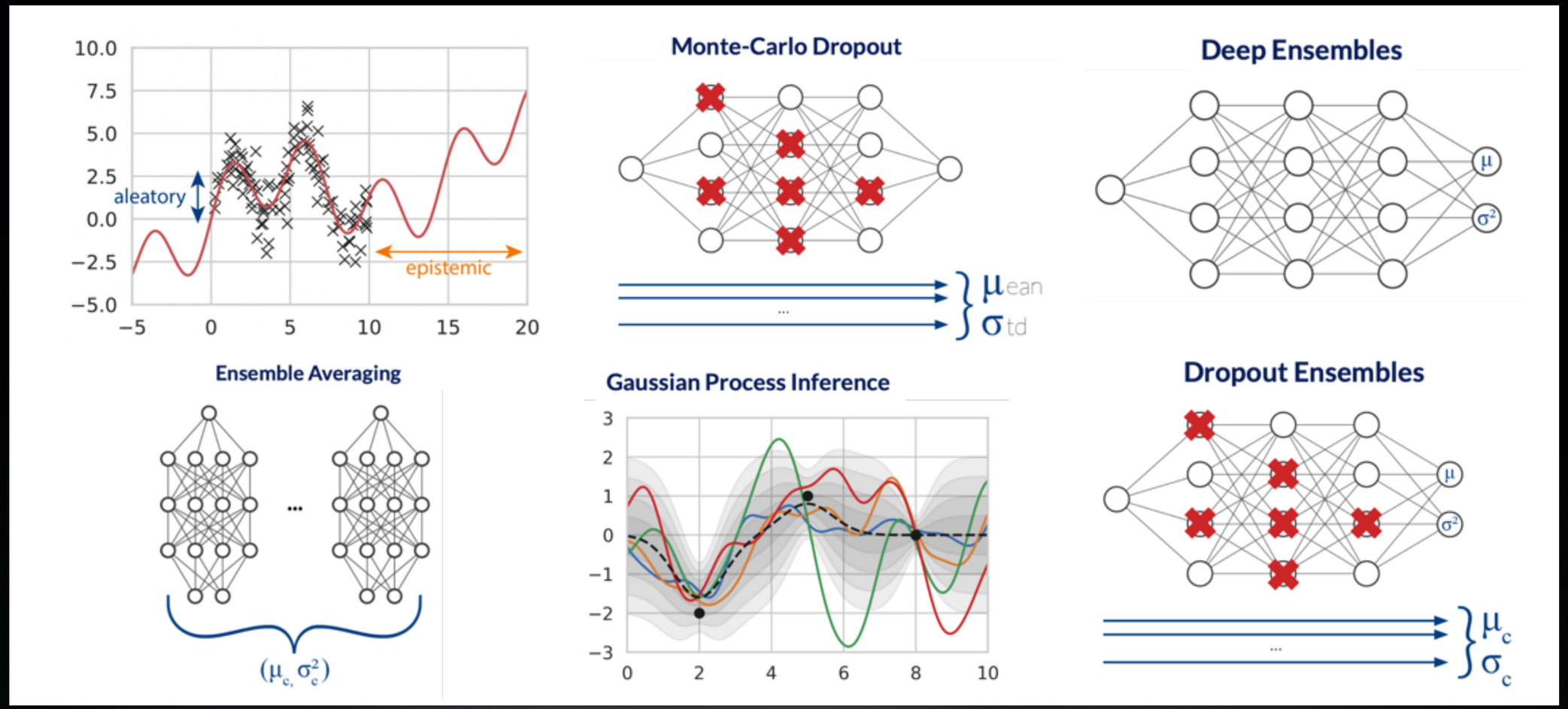
How certain is the prediction?



https://www.researchgate.net/figure/Illustration-of-Gaussian-process-regression-in-one-dimension-for-the-target-test_fig1_327613136

UNCERTAINTY QUANTIFICATION

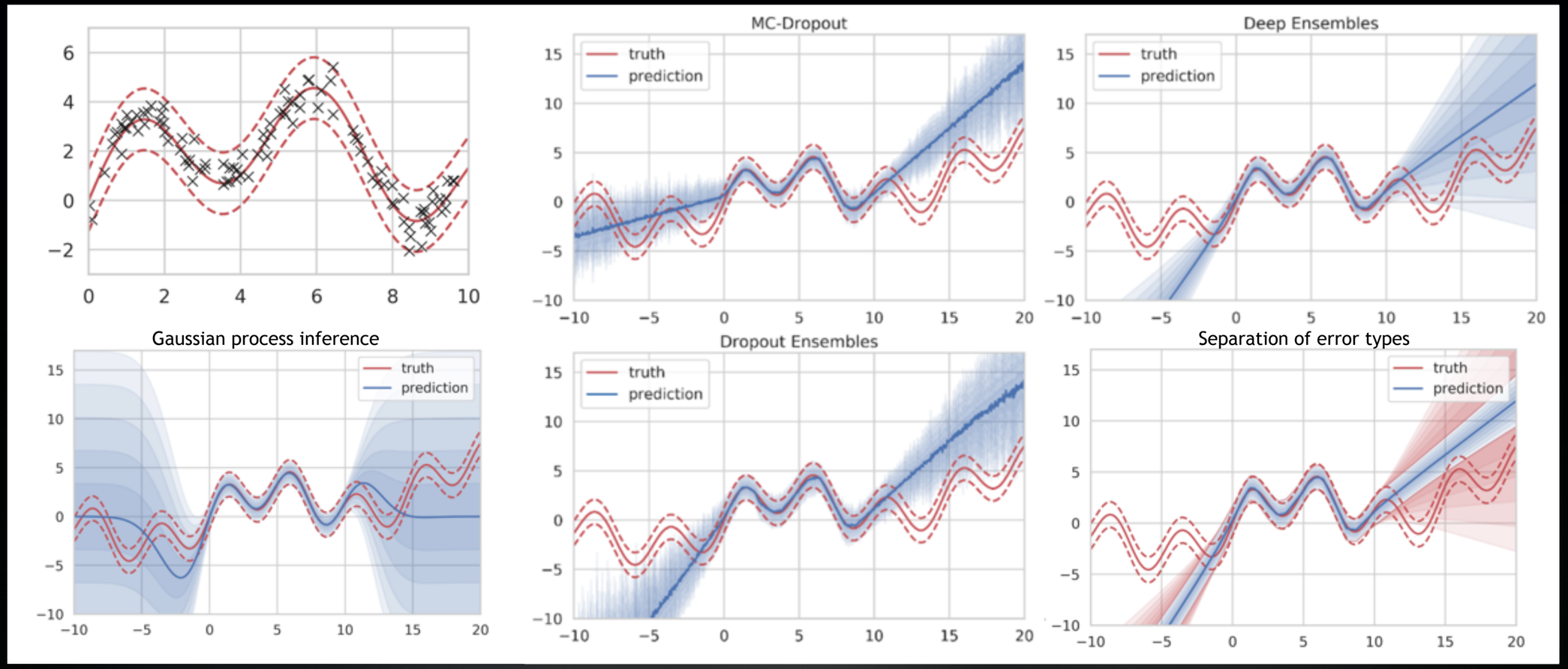
Methods for quantifying uncertainty



https://www.inovex.de/blog/uncertainty-quantification-deep-learning/

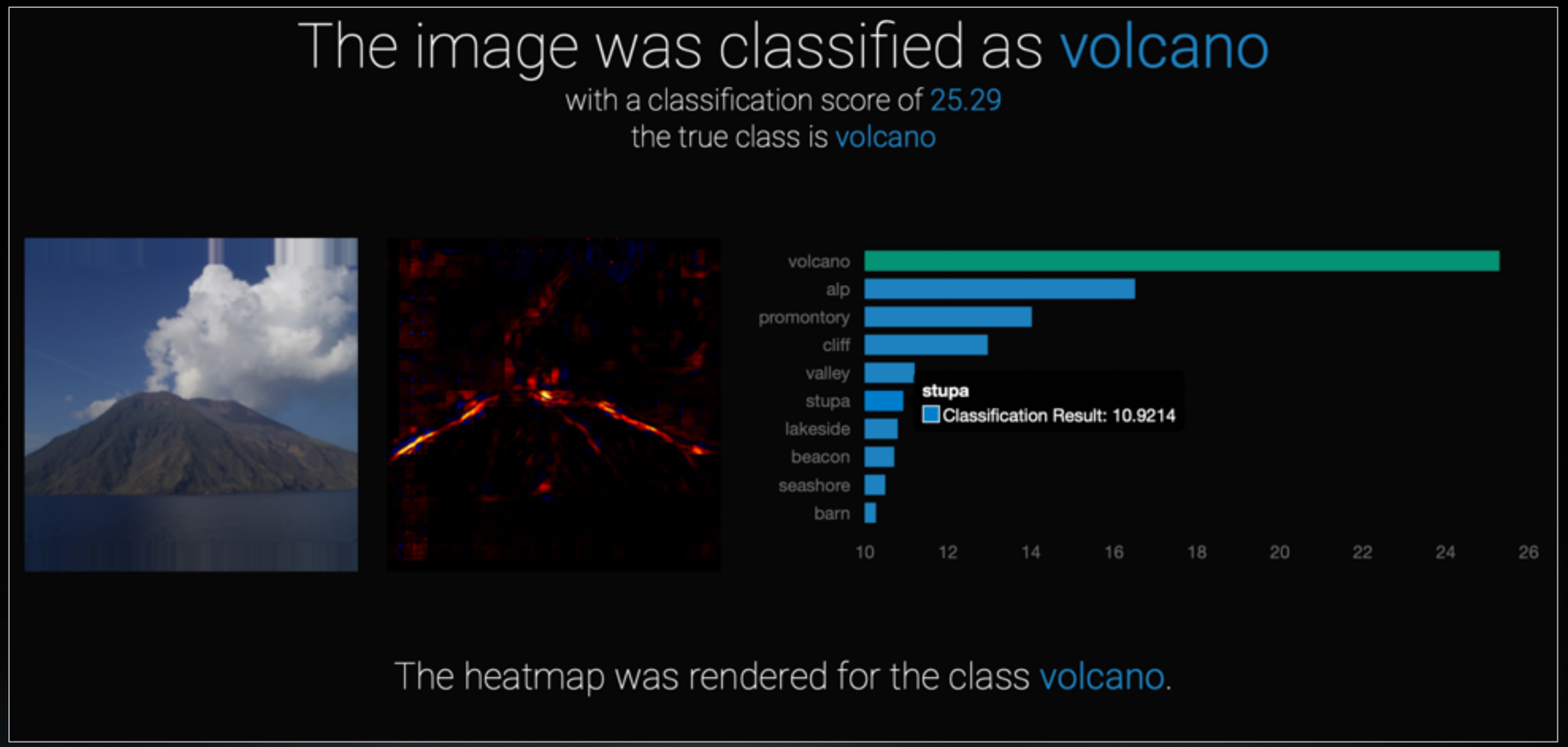
UNCERTAINTY QUANTIFICATION

Methods for quantifying uncertainty



INTERPRETABILITY

What criteria were used in this prediction?



Layer-wise Relevance Propagation (LRP)

https://lrpserver.hhi.fraunhofer.de/image-classification

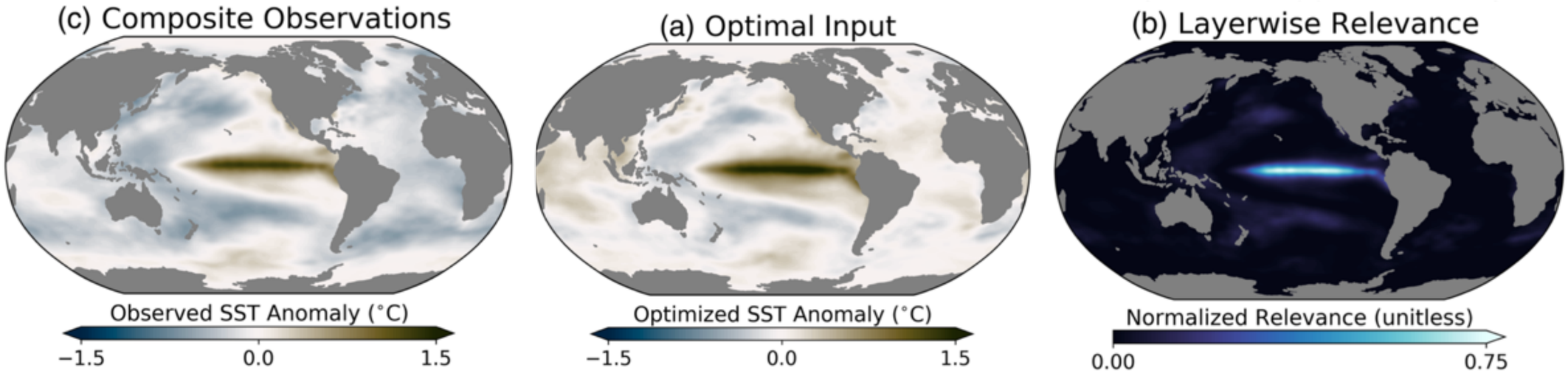
INTERPRETABILITY

Backwards optimization and Layerwise Relevance Propagation applied to ENSO

Physically Interpretable Neural Networks for the Geosciences: Applications to Earth System Variability

Benjamin A. Toms[1], Elizabeth A. Barnes[1], and Imme Ebert-Uphoff[2,3]

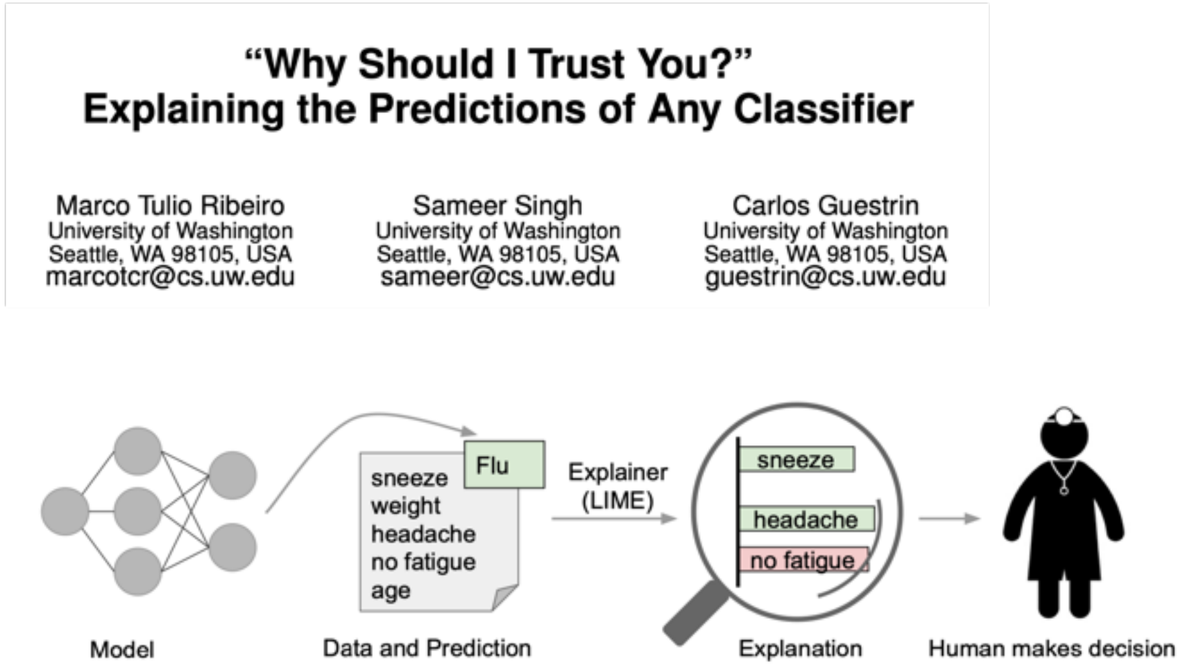
[1]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA, [2]Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO, USA, [3]Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA

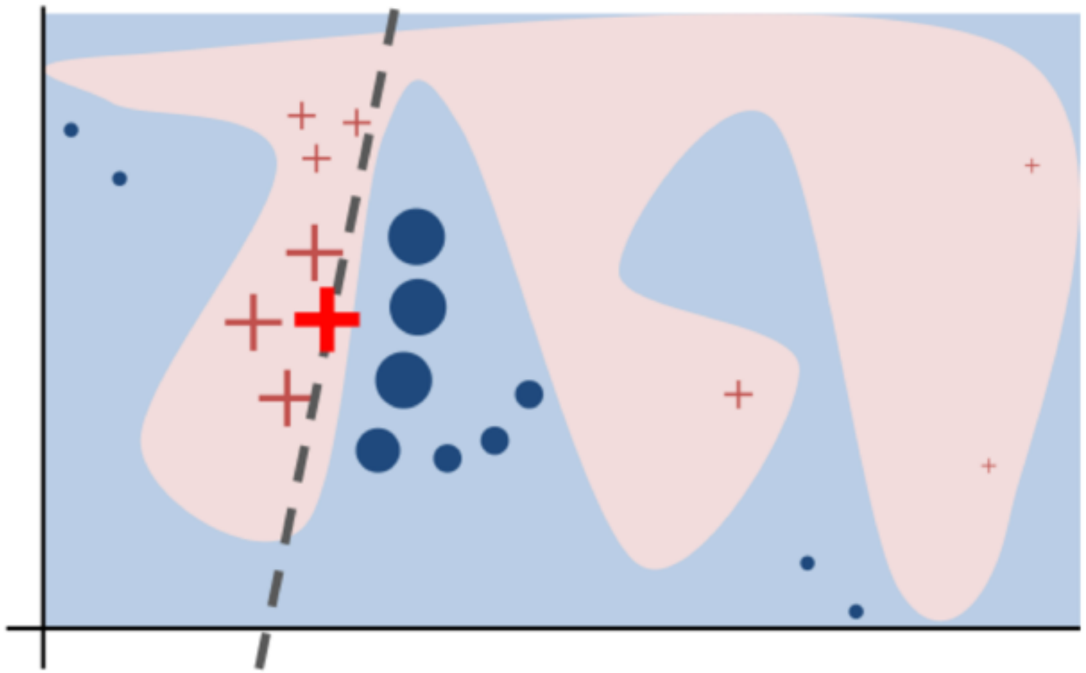


https://doi.org/10.1029/2019MS002002

INTERPRETABILITY

LIME: Local interpretable model-agnostic explanations





https://arxiv.org/pdf/1602.04938.pdf

FORTRAN / AI COUPLING

How can I glue my AI and HPC code together?

+

FORTRAN / AI COUPLING

Many solutions. None ideal.

Use Julia. Call Fortran

julia

flux

DiffEqFlux.jl

GPUArrays

Use C++ Instead

mxnet

8 Language Bindings

Deep integration into Python and support for Scala, Julia, Clojure, Java, C++, R and Perl.

GluonCV

GluonNLP

GluonTS

Ok, but limited

Keras

The Bridge

Fortran

Implemented layers

- Dense
- Dropout
- Batch Normalization

Missing: Native API

FORTRAN

INPUT: REASON • OUTPUT: PLEASURE

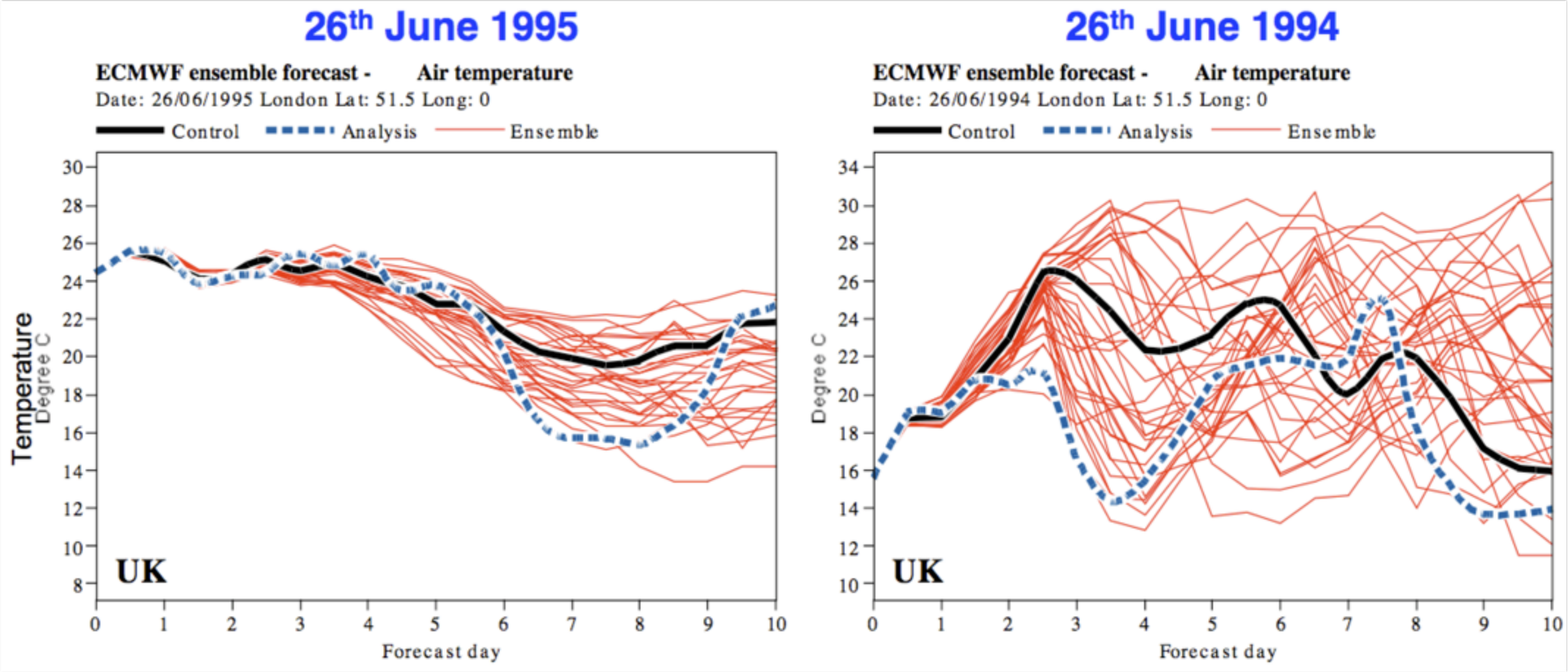
Missing: Fortran Bindings

PYTORCH

TensorFlow

OVERCOMING LOSS OF DYNAMIC RANGE

Ensemble mean always has less variability than the ensemble members



OVERCOMING LOSS OF DYNAMIC RANGE

Stochastic Parameterizations

Machine Learning for Stochastic Parameterization: Generative Adversarial Networks in the Lorenz '96 Model

David John GagneII[1], Hannah M. Christensen[1,2], Aneesh C. Subramanian[3], and Adam H. Monahan[4]

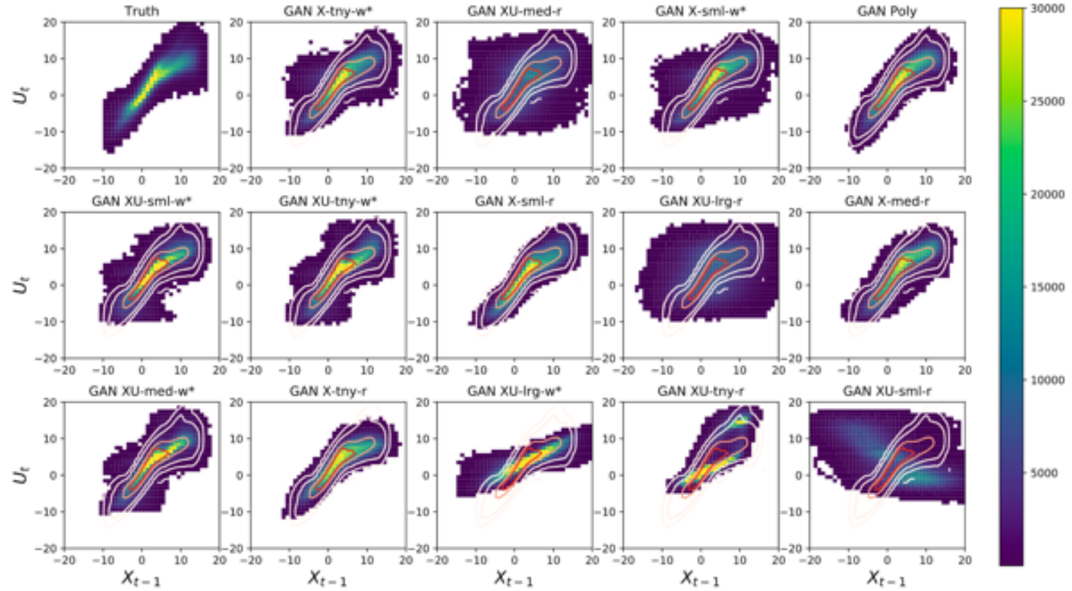
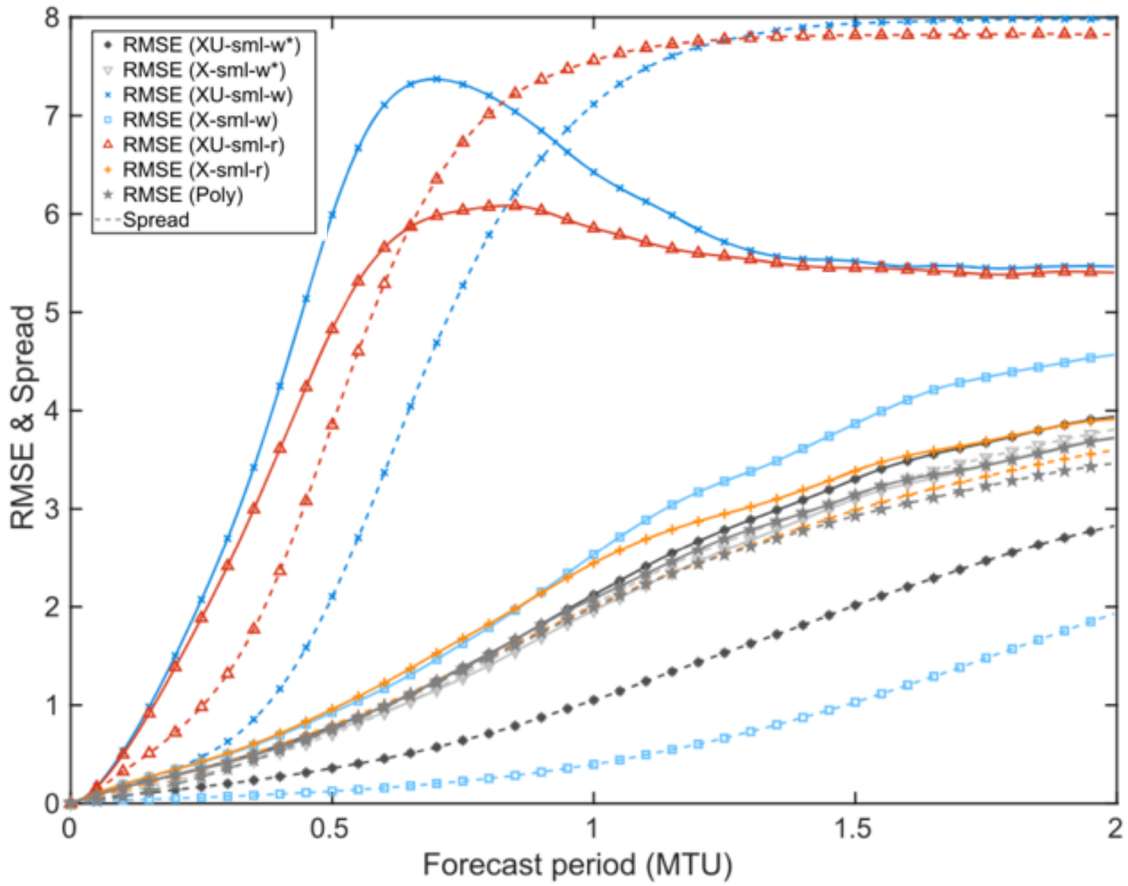


Figure 3. Joint distributions (2-D histograms) of X_{t-1} and U_t for each GAN configuration. The truth joint distribution is overlaid in red contours on each forecast model distribution. The distributions are ordered from left to right descending in terms of their relative total marginal Hellinger distances.



TRENDS AND BREAKTHROUGHS

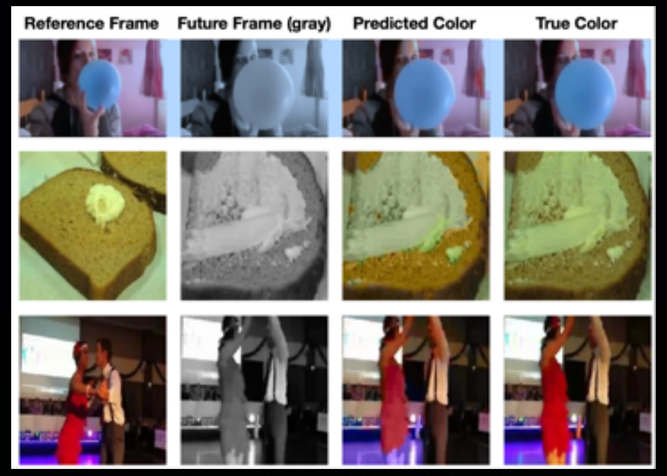
SELF-SUPERVISION

Babies learn about the world without large labelled datasets. AI can too.

SELF-SUPERVISION

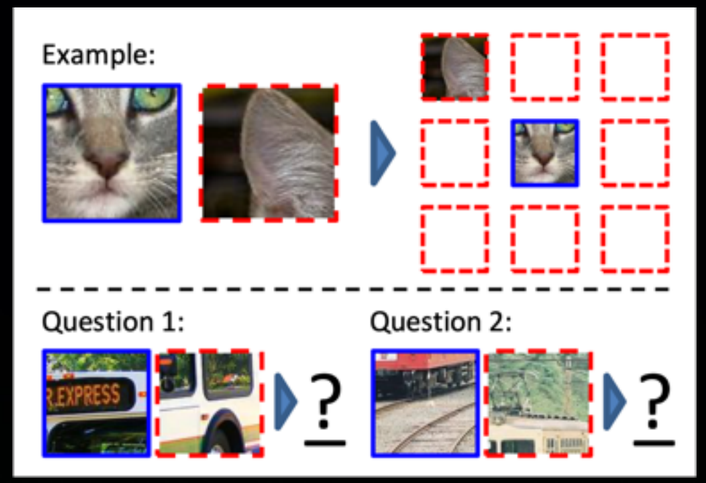
Pretext tasks build up an internal representation

PREDICT COLOR

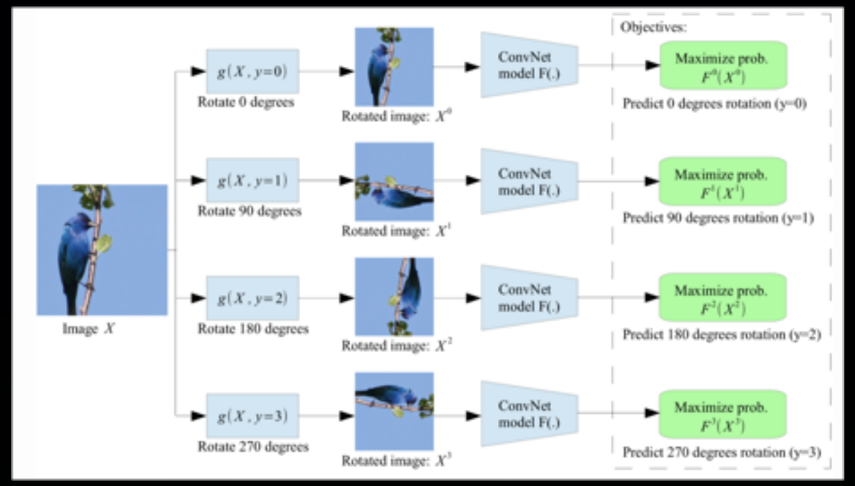


https://arxiv.org/pdf/1806.09594.pdf

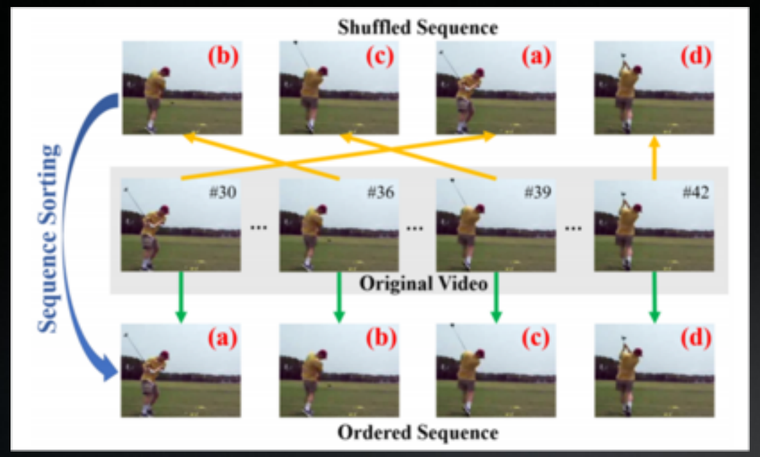
PREDICT SPATIAL RELATIONSHIPS



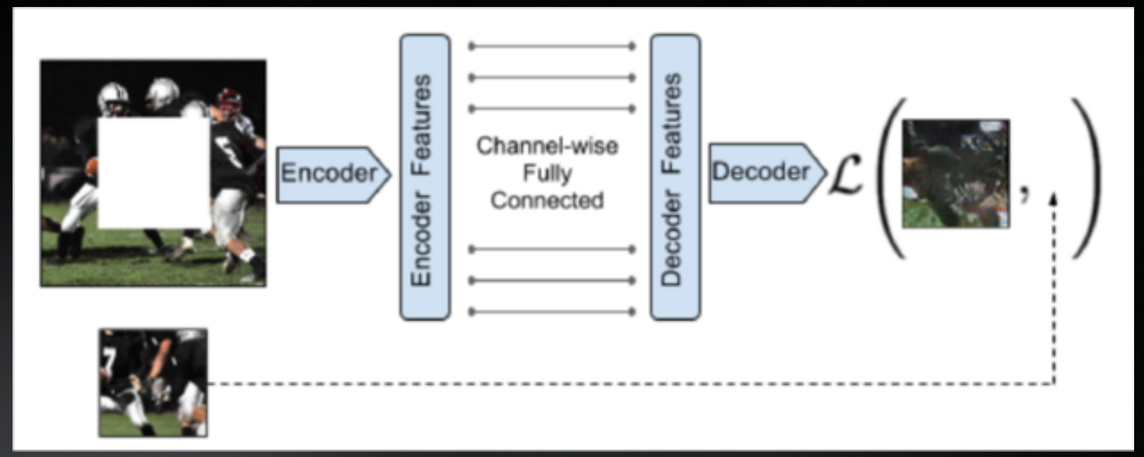
PREDICT ORIENTATION



PREDICT TEMPORAL ORDER

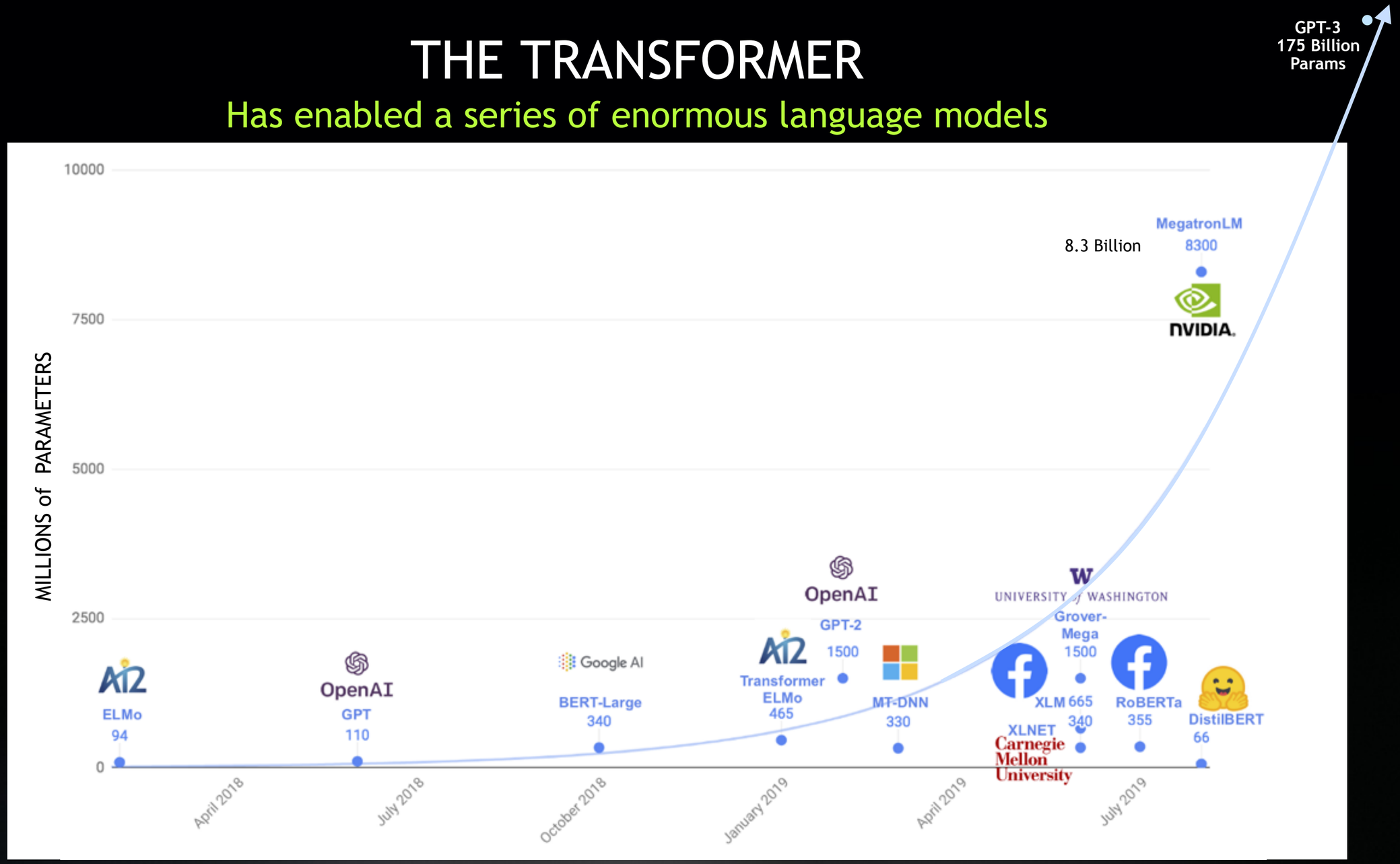


PREDICT MISSING PIECES



THE TRANSFORMER

Has enabled a series of enormous language models



best language papers encoder-decoder-transformers

GPT-3

(July 2020) Current king of the languages models. Based upon the transformer

Language Models are Few-Shot Learners

The three settings we explore for in-context learning

Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

```
Translate English to French:      ← task description
cheese =>                         ← prompt
```

One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.

```
Translate English to French:      ← task description
sea otter => loutre de mer        ← example
cheese =>                         ← prompt
```

Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

```
Translate English to French:      ← task description
sea otter => loutre de mer        ← examples
peppermint => menthe poivrée
plush girafe => girafe peluche
cheese =>                         ← prompt
```

What is the average number of influencers each user is subscribed to?

```
SELECT
  avg(count)
FROM
  (
    SELECT
      user_id,
      count(*)
    FROM
      subscribers
    GROUP BY
      user_id
```

Equation description

integral from a to b of f(t) with respect to t = F of b minus F of a

Translate

$$\int_a^b f(t)\,dt = \int_a^b \frac{F(b)-F(a)}{t}\,dt$$

How many carbon atoms are there in Benzene?

There are 6 carbon atoms in Benzene.

More info on this / Show more results

large text that says "WELCOME TO MY NEWSLETTER" and a blue button that sa

Generate

```
<h1 style={{fontSize: 50, color: 'white'}}>WELCOME TO MY NEWSLETTER</h1><button
style={{color: 'white', backgroundColor: 'blue'}}>Subscribe</button>
```

WELCOME TO MY NEWSLETTER

Subscribe

Build Keras Models

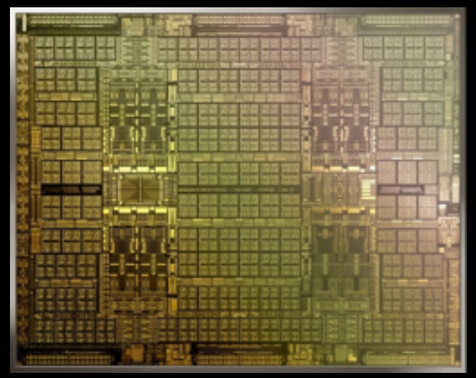
Build a model to classify images into 5 groups. The dataset has 25000 images, with an input shape of 500x500.

Generate Model

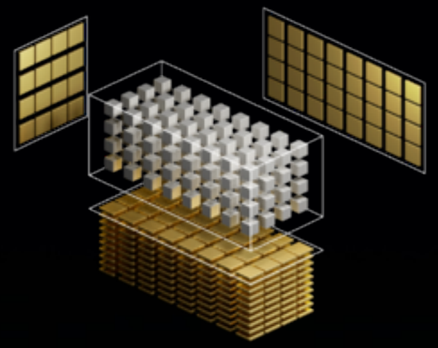
```
from keras.models import Sequential from
keras.layers import Conv2D, MaxPooling2D,
Dropout, Flatten, Dense, Activation,
BatchNormalization model = Sequential()
model.add(Conv2D(32, (5,5)
```

NEW HARDWARE AND SOFTWARE TOOLS

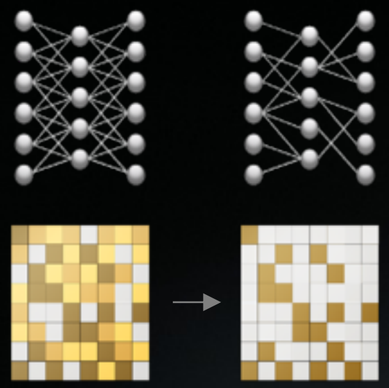
5 MIRACLES OF A100



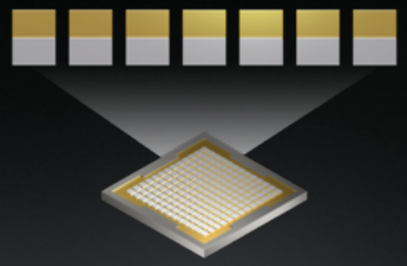
NVIDIA Ampere Architecture
World's Largest 7nm chip
54B XTORS, HBM2



3rd Gen Tensor Cores
Faster, Flexible, Easier to use
20x AI Perf with TF32
2.5x HPC Perf



New Sparsity Acceleration
Harness Sparsity in AI Models
2x AI Performance



New Multi-Instance GPU
Optimal utilization with right sized GPU
7x Simultaneous Instances per GPU

3rd Gen NVLINK and NVSWITCH
Efficient Scaling to Enable Super GPU
2X More Bandwidth

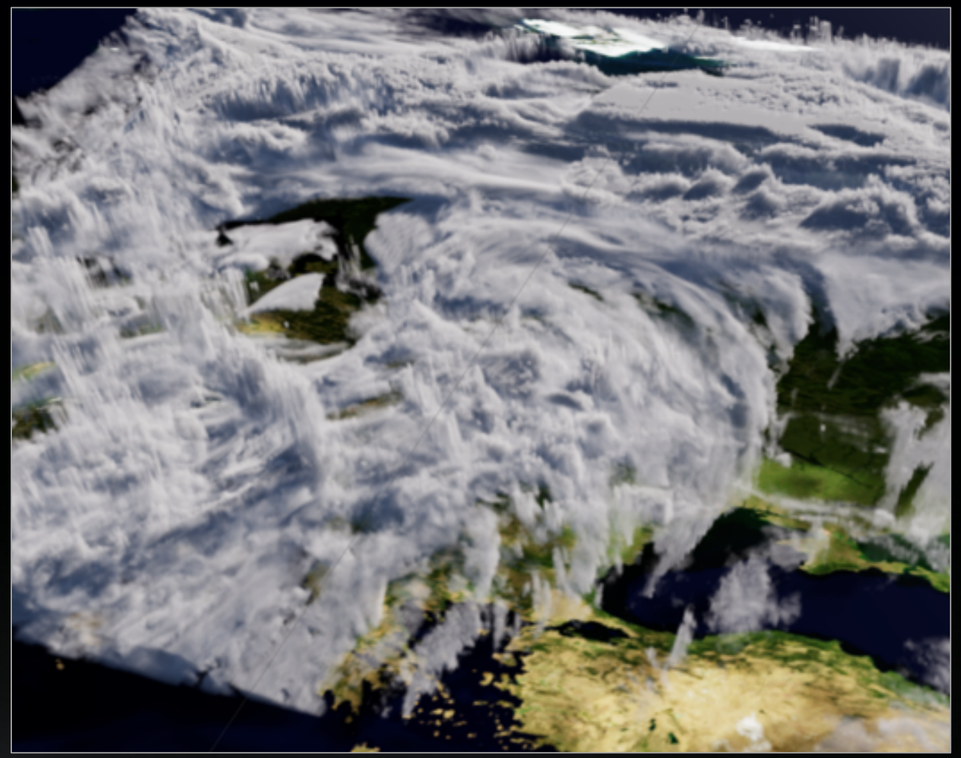
OMNIVERSE

Interactive Raytracing for Data Visualization And Graphical User Interfaces

https://developer.nvidia.com/nvidia-omniverse-platform

OMNIVERSE

Interactive Raytracing for Data Visualization And Graphical User Interfaces



PYTORCH LIGHTNING

API to standardize and accelerate your PyTorch models

Common Use Cases

- 16-bit training
- Computing cluster (SLURM)
- Child Modules
- Debugging
- Experiment Logging
- Experiment Reporting
- Early stopping
- Fast Training
- Model Hooks
- Hyperparameters
- Learning Rate Finder
- Multi-GPU training
- Multiple Datasets
- Saving and loading weights
- Optimization
- Performance and Bottleneck Profiler
- Single GPU Training
- Sequential Data
- Training Tricks
- Transfer Learning
- TPU support
- Test set
- Inference in Production

https://reproducibility-challenge.github.io/neurips2019/resources/

PYTORCH-LIGHTNING BOLTS

Lightning implementations of popular models, optimized for GPUs

PyTorch Lightning Bolts

Models and model components

```
from pl_bolts.models import VAE, LogisticRegression, ImageGPT
from pl_bolts.models.resnets import resnet50
```

SOTA models

```
from pl_bolts.models.self_supervised import SimCLR
from pl_bolts.models.self_supervised import SwAV
```

Callbacks

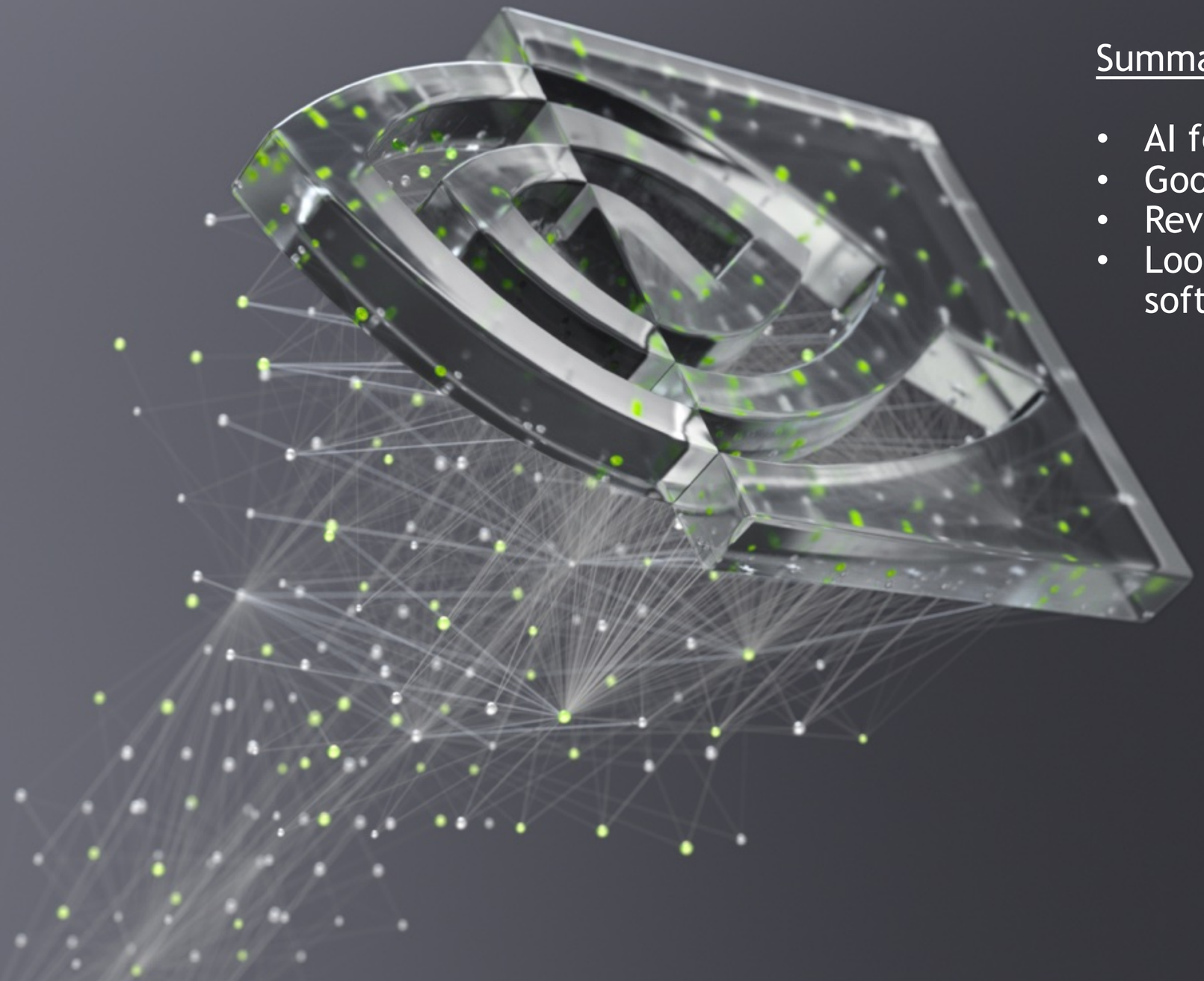
```
from pl_bolts.callbacks.self_supervised import SSLOnlineEvaluator
```

DataModules

```
from pl_bolts.datamodules import CIFAR10DataModule, FashionMNISTDataModule
```

Mix and match your data and lightning code

```
model = ImageGPT(datamodule=FashionMNISTDataModule(PATH))
```



Summary

- AI for science is advancing rapidly!
- Good progress on scientific challenges
- Reviewed major trends in AI
- Looked at powerful new hardware and software tools you can use

dhall@NVIDIA