# The deep learning revolution in weather forecasting



David Landry, PhD Candidate @ INRIA ML for Climate Meetup 2023-09-21

### About me

- PhD Candidate @ INRIA under Claire
  Monteleoni
  - Recently created <u>ARCHES</u> team
- Previously: Data Scientist, notably @ the Meteorological Service of Canada
- Computer science perspective



# The quiet revolution

- Numerical weather prediction (NWP) has been steadily progressing over the past decades
- Great scientific achievement
  - Complex simulation problem being Ο solved **routinely** in operational forecasting centers across the world

#### Strong socio-economical value

Climate hazards mitigation, Logistics, 0 Property loss prevention, ...

#### REVIEW

doi:10.1038/nature14956

#### The quiet revolution of numerical weather prediction

Peter Bauer<sup>1</sup>, Alan Thorpe<sup>1</sup> & Gilbert Brunet<sup>2</sup>

Advances in numerical weather prediction represent a quiet revolution because they have resulted from a steady accumulation of scientific knowledge and technological advances over many years that, with only a few exceptions, have not been associated with the aura of fundamental physics breakthroughs. Nonetheless, the impact of numerical weather prediction is among the greatest of any area of physical science. As a computational problem, global weather prediction is comparable to the simulation of the human brain and of the evolution of the early Universe, and it is performed every day at major operational centres across the world.

t the turn of the twentieth century, Abbe1 and Bjerknes2 proposed that the laws of physics could be used to forecast the weather; they recognized that predicting the state of the atmosphere could be treated as an initial value problem of mathematical physics, wherein future weather is determined by integrating the governing partial differential equations, starting from the observed current weather. This proposition, even with the most optimistic interpretation of Newtonian determinism, is all the more audacious given that, at that time, there were few routine observations of the state of the atmosphere, no computers, and little understanding of whether the weather possesses any significant degree of predictability. But today, more than 100 years later, this paradigm translates into solving daily a system of nonlinear differential equations at about half a billion points per time step between the initial time and weeks to months ahead, and accounting for dynamic, thermodynamic, radiative and chemical processes working on scales from hundreds of metres to thousands of kilometres and from seconds to weeks.

A touchstone of scientific knowledge and understanding is the ability to predict accurately the outcome of an experiment. In meteorology, this translates into the accuracy of the weather forecast. In addition, today's numerical weather predictions also enable the forecaster to assess quantitatively the degree of confidence users should have in any particular forecast. This is a story of profound and fundamental scientific success built upon the application of the classical laws of physics. Clearly the success has required technological acumen as well as scientific advances and vision.

Accurate forecasts save lives, support emergency management and mitigation of impacts and prevent economic losses from high-impact weather, and they create substantial financial revenue-for example, in energy, agriculture, transport and recreational sectors. Their substantial benefits far outweigh the costs of investing in the essential scientific research, super-computing facilities and satellite and other observational programmes that are needed to produce such forecasts3.

These scientific and technological developments have led to increasing weather forecast skill over the past 40 years. Importantly, this skill can be objectively and quantitatively assessed, as every day we compare the forecast with what actually occurs. For example, forecast skill in the range from 3 to 10 days ahead has been increasing by about one day per decade: today's 6-day forecast is as accurate as the 5-day forecast ten years ago, as shown in Fig. 1. Predictive skill in the Northern and Southern hemispheres is almost equal today, thanks to the effective

use of observational information from satellite data providing global coverage

More visible to society, however, are extreme events. The unusual path and intensification of hurricane Sandy in October 2012 was predicted 8 days ahead, the 2010 Russian heat-wave and the 2013 US cold spell were forecast with 1-2 weeks lead time, and tropical sea surface temperature variability following the El Niño/Southern Oscillation phenomenon can be predicted 3-4 months ahead. Weather and climate prediction skill are intimately linked, because accurate climate prediction needs a good representation of weather phenomena and their statistics, as the underlying physical laws apply to all prediction time ranges.

This Review explains the fundamental scientific basis of numerical weather prediction (NWP) before highlighting three areas from which the largest benefit in predictive skill has been obtained in the pastphysical process representation, ensemble forecasting and model initialization. These are also the areas that present the most challenging science questions in the next decade, but the vision of running



Figure 1 A measure of forecast skill at three-, five-, seven- and ten-day ranges, computed over the extra-tropical northern and southern hemispheres. Forecast skill is the correlation between the forecasts and the verifying analysis of the height of the 500-hPa level, expressed as the anomaly with respect to the climatological height. Values greater than 60% indicate useful forecasts, while those greater than 80% represent a high degree of accuracy. The convergence of the curves for Northern Hemisphere (NH) and Southern Hemisphere (SH) after 1999 indicates the breakthrough in exploiting satellite data through the use of variational data<sup>100</sup>

#### doi:10.1038/nature14956



#### **ECMWF Global Forecast performance over 4 decades**

https://charts.ecmwf.int/products/plwww\_m\_hr\_ccaf\_adrian\_ts?single\_product=latest

# The problem ahead

Some/most of the progress had been related to Moore's law

Growth in CPU computing power slowed down



### What happened

- Starting in 2023, three different research teams used deep neural networks to perform medium-range weather forecasts
- Their skill became ~equivalent to the leading NWP models

Forecasting	g Global	Weather	
ith Graph	Neural	Networks	

Pron Kolder rkeisler@gmail.com

#### Abstract

We present a data-driven approach for forecasting global weather using graph neural networks. The system learns to step forward the current 3D atmospher state by six hours, and multiple steps are chained together to produce skillu forecasts going out several days into the future. The underlying model is trained on translysis data from ERA5 or forecast data from GES. Test performance or on reality's data from ERCS or forecast data from Gr3. Test performance on metrics such as 2500 (geopotential height) and 7850 (temperature) improves upon previous data-driven approaches and is comparable to operational, full-resolution, physical models from GFS and ECMWF, at least when evaluated on 1-degree cales and when using reanalysis initial conditions. We also show results from necting this data-driven model to live, operational forecasts from GFS

#### 1 Introduction

Numerical weather prediction (NWP), as part of the broader weather enterprise, has had an eno and positive intract on society. Decades of steady improvements in the quantity and types of and positive induction society. Declared when any induction of the positive inducting induction of the positive induction of the positive induction

While statistical techniques have been used within NWP for decades, the core dynamical engines of these models excitate events and the provide the physical principles governing the atmosphere and occan. More recently, spurred on by advancements in machine learning (ML), there has been a surge of interest in statistical, data-driven techniques for watcher forecosting. The motivation for using ML is to improve upon an already extremely successful WWP program through some contrained to using the is to improve upon an already extremely successful WWP program through some contraintion of better forecasts, faster forecasts, or more forecasts, i.e. larger ensembles. There may also be opportunities for using ML to advance our scientific understanding of the underlying physical processes [Cranme] erail [020].

There is currently a very active bub of research at the intersection of NWP and M  FOURCASTNET: A GLOBAL DATA-DRIVEN HIGH-RESOLUTION WEATHER MODEL USING ADAPTIVE FOURIER NEURAL **OPERATORS** 

	AFREEKIST	
Jaideep Pathak NVIDIA Corporation Santa Clara, CA 95051	Shashank Subramanian Peter Lawrence Berkeley Lawre National Laboratory Nation Berkeley, CA 94720 Berkel	Harrington Sanjeev Raja nee Berkeley University of Michiga al Laboratory Ann Arbor, MI 48105 ey, CA 94720
Ashesh Chattopadh Rice University Houston, TX 770	yay Morteza Mardani NVIDIA Corporation 5 Santa Clara, CA 95051	Thorsten Kurth NVIDIA Corporation Santa Clara, CA 95051
David Hall NVIDIA Corporation Santa Clara, CA 95051	Zongyi Li California Institute of Technolo Pasadena, CA 91125 NVIDIA Corporation Santa Clara, CA 95051	Kamyar Azizzadenesheli pgy Purdue University West Lafayette, IN 47907
Pedram Hassanzadeh Rice University Houston, TX 77005	Karthik Kashinath NVIDIA Corporation Santa Clara, CA 95051	Animashree Anandkumar California Institute of Technology Pasadena, CA 91125 NVIDIA Corporation Santa Clara, CA 95051
	February 24, 2022	
	ABSTRACT	

2202.11214v

energy resources, predicting extreme weather events such as tropical cyclones, extra-tropical cyclone and atmospheric rivers. FourCastNet matches the forecasting accuracy of the ECMWF Integrated Forecasting System (IFS), a state-of-the-art Numerical Weather Prediction (NWP) model, at short lead times for large-scale variables, while outperforming IFS for small-scale variables, including precipita tion. FourCastNet renerates a week-long forecast in less than 2 seconds, orders of magnitude faste than IFS. The speed of FourCastNet enables the creation of rapid and inexpensive large-

#### TECHNICAL REPORT

#### Pangu-Weather: A 3D High-Resolution System for Fast and Accurate Global Weather Forecast

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian<sup>22</sup>, Fellow, IEEE

Abstract—In this paper, we present Parqu-Weather, a deep learning based system for fast and acquirate clobal weather forecast. For Automation in the paper, we present any prevention, a breep reaming basics system on ass and activity global weather data from the Sh paration of this purpose, we establish a data driven environment by downloading 41 years of hourly global weather data from the Sh paration of ECMWF reanitysis (ERAS) data and train a lew deep neural networks with about 20% million parameters in total. The spatial resolution of forecast is 0.25" × 0.25", comparable to the EGMWF Integrated Forecast Systems (IFS), More importantly, for the first time, an Al-based method outperforms state-of-the-art numerical weather prediction (NWP) methods in terms of accuracy (latitude-weighter RMSE and ACC) of all factors (e.g., geopotential, specific humidity, wind speed, temperature, etc.) and in all time ranges (from one hour to one week). There are two key strategies to improve the prediction accuracy: (i) designing a 3D Farth Specific Transformer Not no one week, interfaite workey stategies or improve the previous accuracy, in oresping a so-care stategies or improve the previous accuracy, in oresping a so-care stategies and a so-(30EST) and the curve of the source of short to medium-range forecast (i.e., forecast time ranges from one hour to one week). Pangu-Weather supports a wide range of downstream forecast scenarios, including extreme weather forecast (e.g., tropical cyclone tracking) and large-member ens forecast in real-time. Pargu-Weather not only ends the debate on whether Al-based methods can surpass conventional NV but also reveals novel directions for improving deep learning weather forecast systems.

Index Terms-Numerical Weather Prediction, Deep Learning, Medium-range Weather Forecas

— • —

#### 

8

Weather forecast is one of the most important scenarios - learning. The methodology is to use a deep neural network of scientific computing. It offers the ability of predicting to capture the relationship between the input (observed future weather changes, especially the occurrence of es-treme weather events (e.g., floods, droughts, hurricanes, cialized computational device (e.g., GPUs), AI-based methetc.), which has large values to the society (e.g., daily activ- ods run very fast and easily achieve a tradeoff between ity agriculture, energy production, transportation, industry, model complexity, prediction resolution, and prediction etc.). In the past decade, with the bloom of high-performance accuracy [9], [10], [11], [12], [13], [14], [15]. As a recent computational device, the community has witnessed a rapid example. FourCastNet [14] increased the spatial resolution development in the research field of numerical weather to  $0.25^{\circ} \times 0.25^{\circ}$ , comparable to the ECMWF Integrated prediction (NWP) []]. Conventional NWP methods mostly follow a simulation-based paradigm which formulates the GPUs for making a 100-member, 2/1-hour forecast, which table equations (PDEs) and solves them using numerical differen-table equations (PDEs) and solves them using numerical methods. However, the forecast accuracy of FourCastNet is simulations 😰, 🚯, 🚯 Due to the high complexity of 🛛 still below satisfaction, e.g., the RMSE of 5-day Z500 forecast solving PDEs, these NWP methods are often very slow, e.g., with a spatial resolution of 0.25° x 0.25°, a single simulation and 462.5, respectively, which are much worse than 333.7 reprocedure for 10-day forecast can take hours of compu-tation using hundreds of nodes in a supercomputer [5]. conjectured that 'a number of fundamental breakthroughs This largely reduces the timeliness in daily weather forecast are needed' before AI-based methods can beat NWP. and the number of ensemble members that can be used The breakthrough comes much earlier than they thought. for probabilistic weather forecast. In addition-

nal In this paper, we present Paper-Weather, a powerfu

#### GraphCast: Learning skillful medium-range global weather forecasting

Remi Lam<sup>7,1</sup>, Alvaro Sanchez-Gonzalez<sup>7,1</sup>, Matthew Willson<sup>7,1</sup>, Peter Wirnsberger<sup>7,1</sup>, Meire Fortunato<sup>7</sup> Perran Alet<sup>\*,1</sup>, Suman Ravuri<sup>\*,1</sup>, Timo Ewalds<sup>1</sup>, Zach Eaton-Rosen<sup>1</sup>, Weihua Hu<sup>1</sup>, Alexander Merose Stephan Hoyer<sup>2</sup>, George Holland<sup>1</sup>, Oriol Vinyals<sup>1</sup>, Jacklynn Stott<sup>1</sup>, Alexander Pritzel<sup>1</sup>, Shakir Mohamed<sup>1</sup> and Peter Battaglia<sup>1</sup> equal contribution, 1Google DeepMind, 2Google Research

Global medium-range weather forecasting is critical to decision-making across many social and eco domains. Traditional numerical weather prediction uses increased compute resources to improve forecast accuracy, but cannot directly use historical weather data to improve the underlying model. We introduce a machine learning-based method called "GraphCast", which can be trained directly from reanalysis data. It predicts hundreds of weather variables, over 10 days at 0.25° resolution slobally, in under one minute. We show that GraphCast significantly outperforms the most accurate operational deterministic systems on 90% of 1380 verification targets, and its forecasts support better severe event prediction including tropical cyclones, atmospheric rivers, and extreme temperatures. GraphCast is a key advance in accurate and efficient weather forecasting, and helps realize the promise of machine learning for modeling complex dynamical systems.

Keywords: Weather forecasting, ECMWF, ERA5, HRES, learning simulation, graph neural networks

#### Introduction

 $\sim$ 

It is 05:45 UTC in mid-October, 2022, in Bologna, Italy, and the European Centre for Medium-Range Weather Forecasts (ECMWF)'s new High-Performance Computing Facility has just started operation For the past several hours the Integrated Forecasting System (IFS) has been running sophisticated calculations to forecast Earth's weather over the next days and weeks, and its first predictions have just begun to be disseminated to users. This process repeats every six hours, every day, to supply the world with the most accurate weather forecasts available.

The IFS and modern weather forecasting more generally are triumnhs of science and engineerin The dynamics of weather systems are among the most complex physical phenomena on Earth, and each day, countless decisions made by individuals, industries, and policymakers depend on accurate weather forecasts, from deciding whether to wear a jacket or to flee a dangerous storm. The dominant approach for weather forecasting today is "numerical weather prediction" (NWP), which involves olving the governing equations of weather using supercomputers. The success of NWP lies in the rigorous and ongoing research practices that provide increasingly detailed descriptions of weather phenomena, and how well NWP scales to greater accuracy with greater computational resources [3, 2] As a result, the accuracy of weather forecasts have increased year after year, to the point where the surface temperature, or the path of a hurricane, can be predicted many days ahead-a possibility that was unthinkable even a few decades ago

# How it works

### Data: ERA5 reanalysis dataset

 Reanalysis: incorporating observations into a physically consistent system to create our best guess of the true state of the system



# How it works

#### Model: Deep Neural Networks

- Various architectural strategies
- Fourier Transforms
- Graph Neural Networks
- Collection of models for resolving different time leaps
- Training loss: deterministic



GraphCast architecture

doi:10.48550/arXiv.2212.12794

### **Computational cost**

Model	Training	Inference
Pangu-Weather	16 days on 192 V100 GPUs	24 hours forecast 1xV100 1.4 seconds on
GraphCast	21 days on 32 Cloud TPU V4 devices (~128 GPUs)	10 days forecast 1 Cloud TPU V4 < 60 seconds
NWP (ECMWF HRES) *Higher resolution	-	10 days forecast 1/3/6 hours 11 664 cores HPC cluster

### A revolution?

Deep weather forecasting models represent a >10 000x decrease in compute cost for a critical piece simulation

A lot of scientific fields and industries depend on the output of weather forecasting models: hydrology, agriculture, risk management, transportation, etc.

Some domains may need to completely revise their modeling stacks to fully integrate the benefits

# Let's speculate

### Foundational models

 Fully-coupled applicative models that perform backpropagation all the way to the weather forecasting models

### • From observation to application

 Model can now integrate more observations with more flexibility because they are fully differentiable

### • Edge computing

- Sensory devices can use their own observations to make adapted forecasts themselves, on the edge
- Extremely short-term forecasts
  - Applications for this?

# **Rate of progress**

- The current rate of progress of artificial intelligence is high
- Deep learning provides a lot of possibilities in integrating various sources of information because it is fully differentiable
  - Capable of multi-modality
- Research is now a lot easier. The rate of iteration on modeling hypothesis has increased dramatically.

# What didn't change

- The existing deep weather forecasting models are based on **reanalysis** products, which themselves contain an NWP model
  - Notably, they need a physical model to decide how to incorporate observations
- Great wealth of knowledge in forecast evaluation, ensemble forecasting, high-performance computing, physics... An operational weather forecasting model is much more than its time stepping
- Deep forecasting networks don't do precipitation -- the distribution is problematic
  - GraphCast computes it but does not report

# What didn't change

- Old conversations are resurfacing
  - Spectral vs Grid-based approaches
- The deep forecasts are blurry, but do we want them sharp?
  - Sharpness is important for interpretation using physics



doi:10.48550/arXiv.2307.10128

# **Upcoming challenges**

- Going probabilistic
- Integrating more fields, including the difficult ones
- Transition towards climate models
- Cascades of models: train area specific, high-resolution models, where data is available
- ...
- Begin to scratch the surface of the possible applications

### References

Bauer, Peter, Alan Thorpe, and Gilbert Brunet. "**The Quiet Revolution of Numerical Weather Prediction**." *Nature* 525, no. 7567 (2015): 47–55. <u>https://doi.org/10.1038/nature14956</u>.

Pathak, Jaideep, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, et al. "FourCastNet: A Global Data-Driven High-Resolution Weather Model Using Adaptive Fourier Neural Operators." *arXiv:2202.11214 [Physics]*, February 22, 2022. <u>http://arxiv.org/abs/2202.11214</u>.

Bi, Kaifeng, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. "Pangu-Weather: A 3D High-Resolution Model for Fast and Accurate Global Weather Forecast." arXiv, November 3, 2022. <u>https://doi.org/10.48550/arXiv.2211.02556</u>.

Lam, Remi, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Alexander Pritzel, Suman Ravuri, et al. "GraphCast: Learning Skillful Medium-Range Global Weather Forecasting." arXiv, December 24, 2022. <u>https://doi.org/10.48550/arXiv.2212.12794</u>.