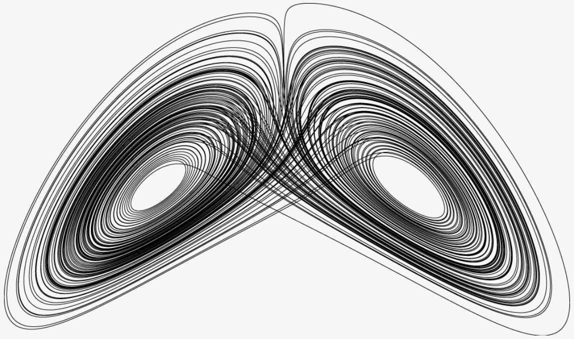
The deep learning revolution in weather forecasting



David Landry, PhD Candidate @ INRIA

ML for Climate Meetup

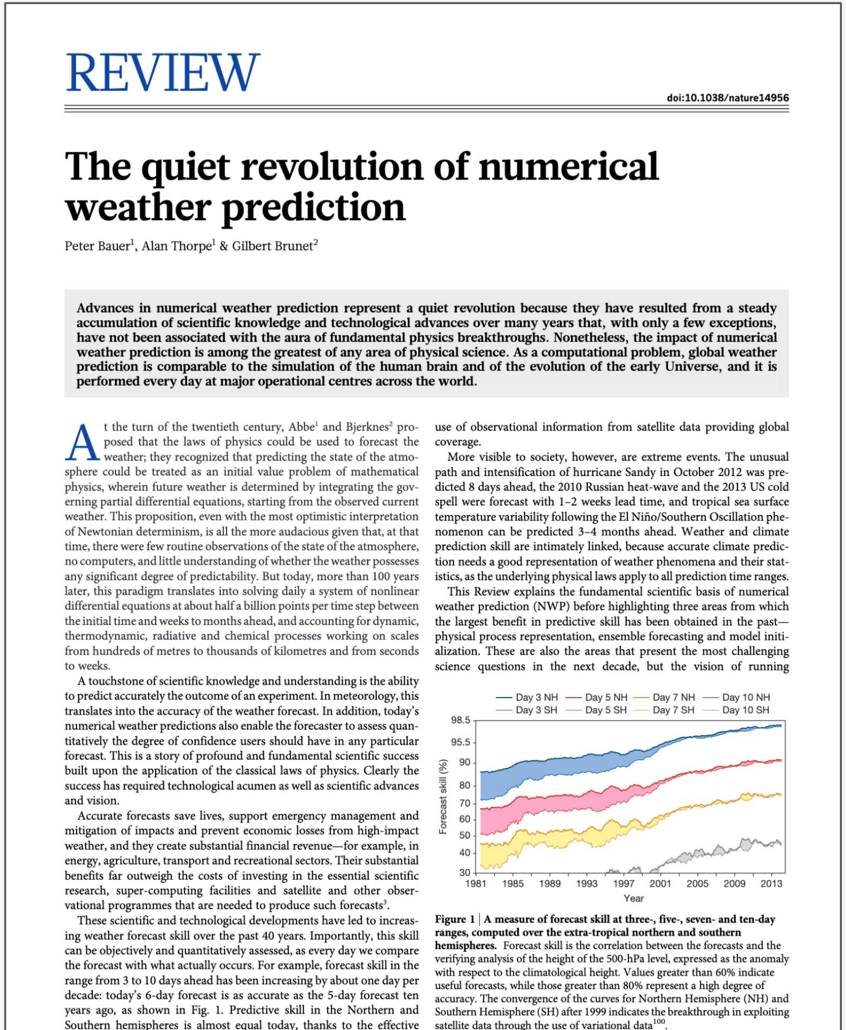
2023-09-21

About me

- PhD Candidate @ INRIA under Claire Monteleoni
    - Recently created ARCHES 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 solved **routinely** in operational forecasting centers across the world
- Strong socio-economical value
  - Climate hazards mitigation, Logistics, Property loss prevention, …


REVIEW

doi:10.1038/nature14956

The quiet revolution of numerical weather prediction

Peter Bauer[1], Alan Thorpe[1] & Gilbert Brunet[2]

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.

At the turn of the twentieth century, Abbe[1] and Bjerknes[2] 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 forecasts[3].

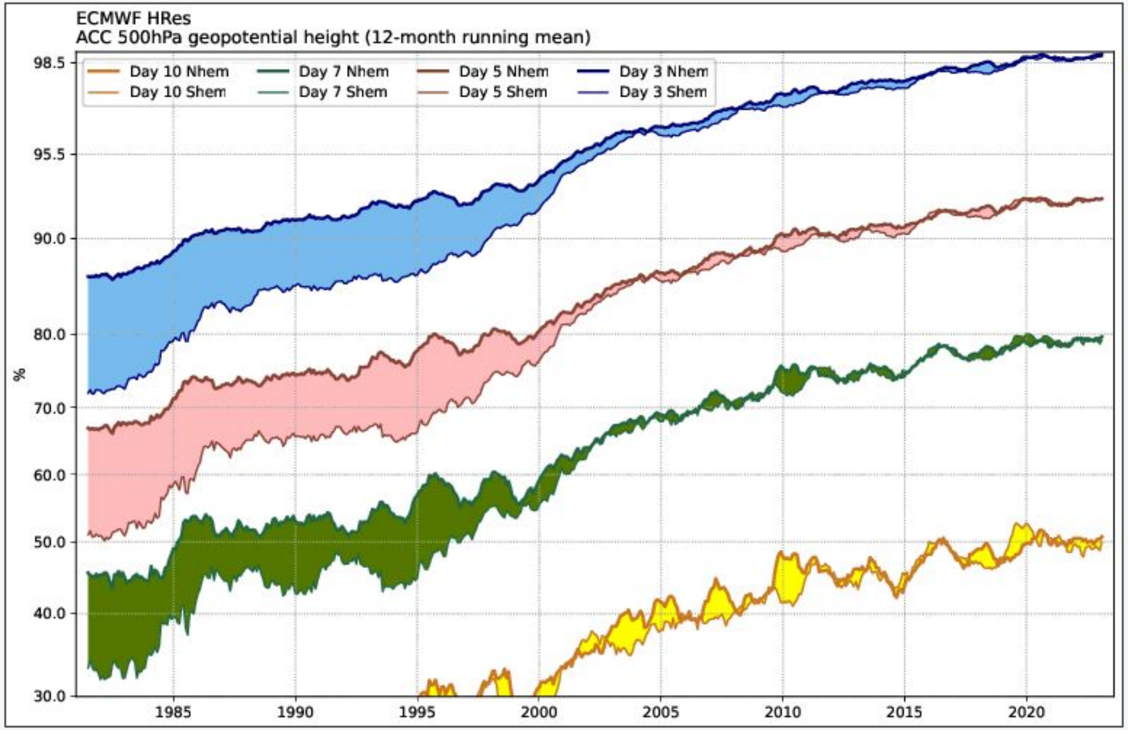
These scientific and technological developments have led to increasing weather forecast skill over the past 40 years. Importantly, this skill can be objectively and quantitatively assessed, as every day we compare the forecast with what actually occurs. For example, forecast skill in the range from 3 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 past—physical 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[100].

doi:10.1038/nature14956



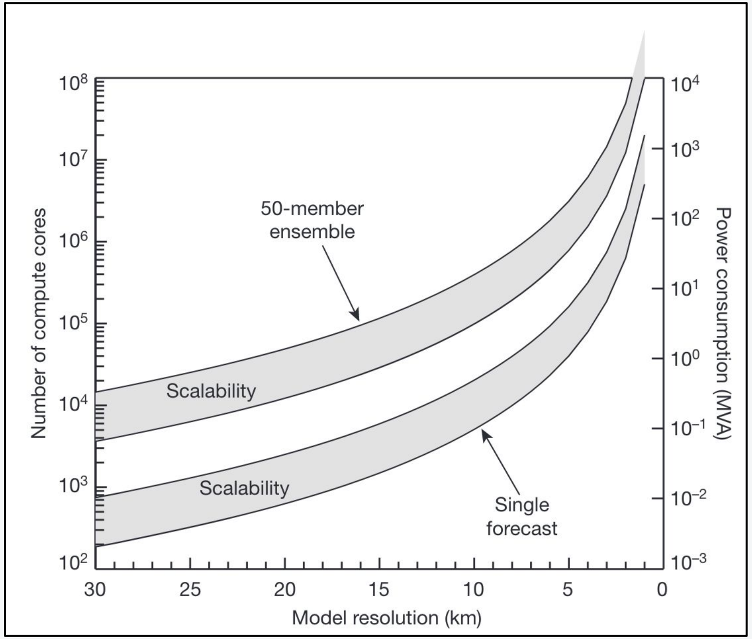
ECMWF Global Forecast performance over 4 decades

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

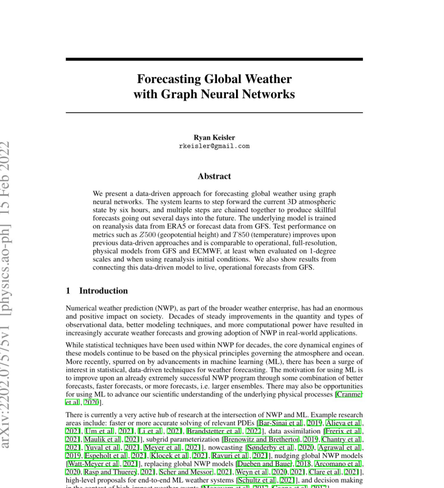
Growth in CPU computing power slowed down

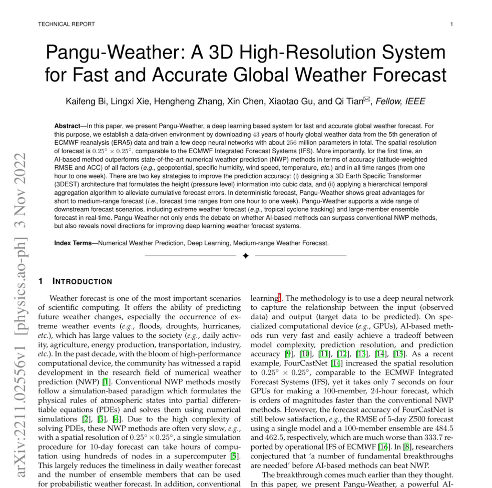


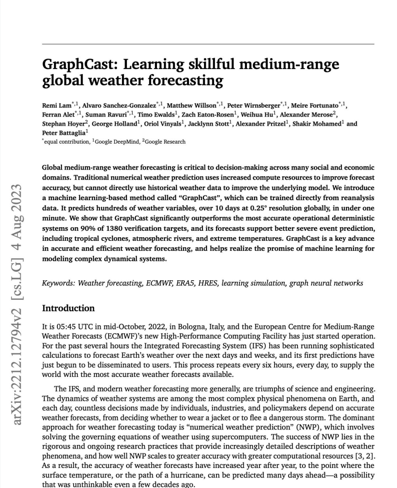
doi:10.1038/nature14956

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



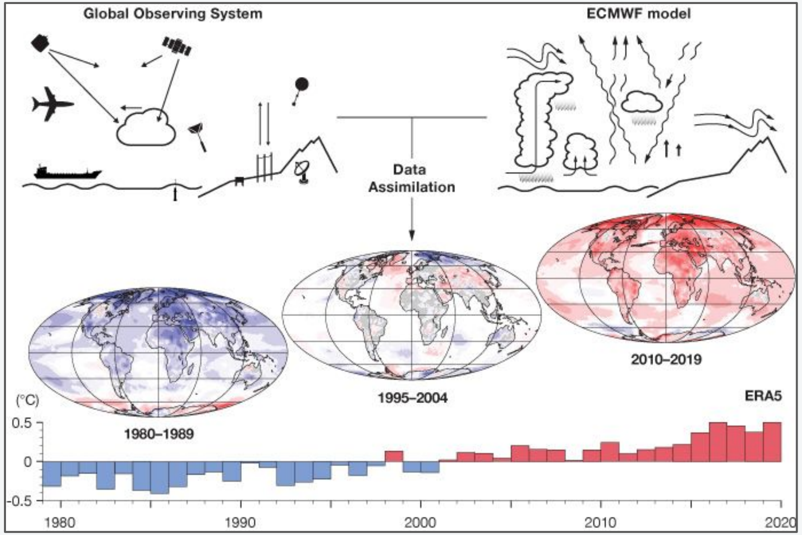




How it works

Data: ERA5 reanalysis dataset

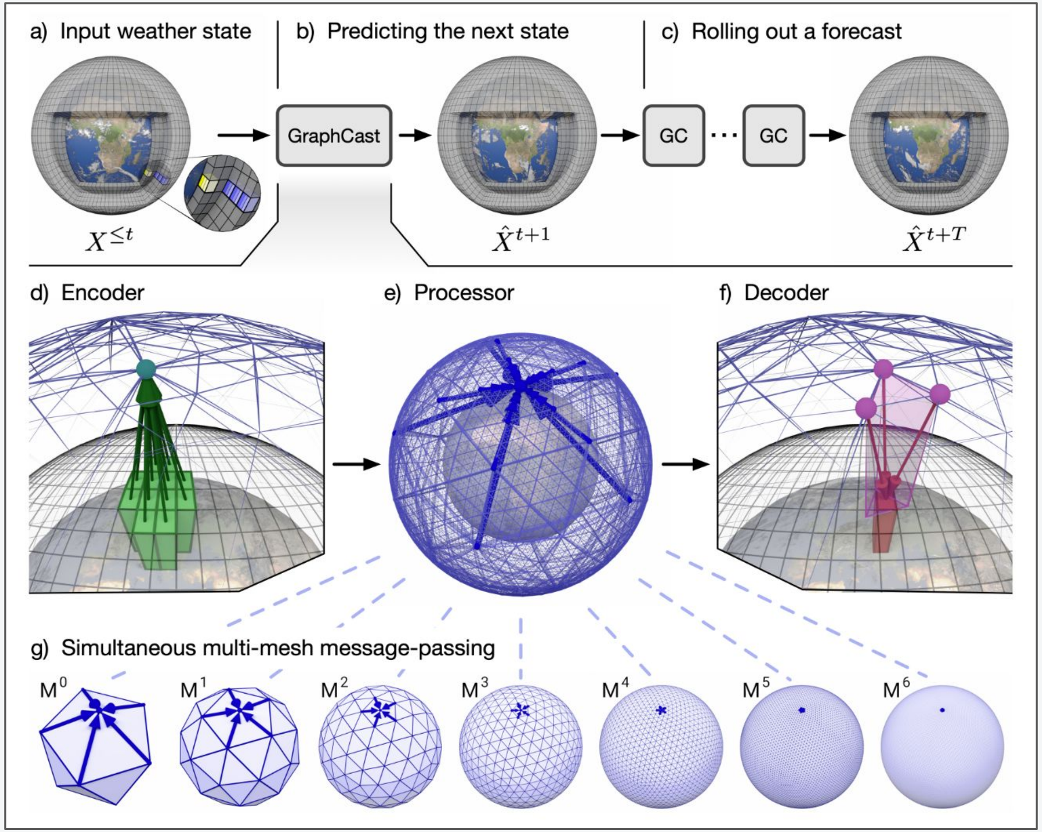
- Reanalysis: incorporating observations into a physically consistent system to create our best guess of the true state of the system



doi:10.1002/qj.3803
https://www.ecmwf.int/en/about/media-centre/focus/2023/fact-sheet-reanalysis

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

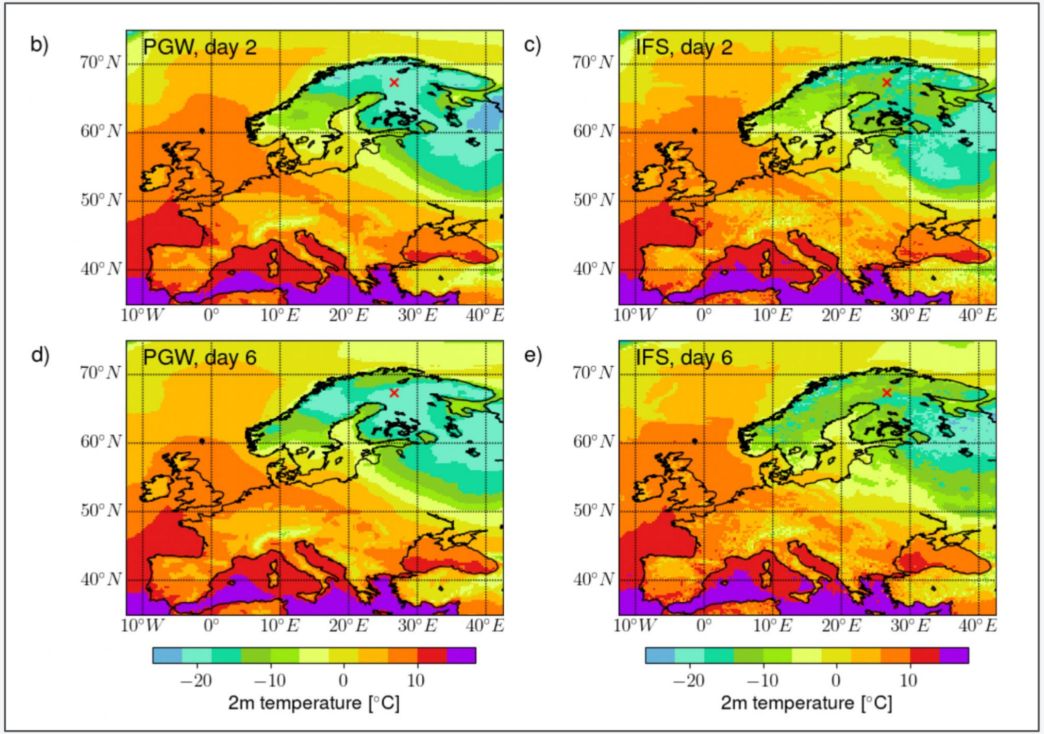
- The current rate of progress of artificial intelligence is high
- Deep learning provides a lot of possibilities in integrating various sources of information because it is fully differentiable
    - Capable of **multi-modality**
- Research is now a lot easier. The rate of iteration on modeling hypothesis 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 <u>didn't</u> 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. https://doi.org/10.1038/nature14956.

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. http://arxiv.org/abs/2202.11214.

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. https://doi.org/10.48550/arXiv.2211.02556.

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. https://doi.org/10.48550/arXiv.2212.12794.

Illustration, title page: https://www.iaacblog.com/programs/algorithmic-emergence-chaotic-attractor-equations/