https://doi.org/10.1038/s43017-023-00468-z

Deep learning and a changing economy in weather and climate prediction

Peter Bauer, Peter Dueben, Matthew Chantry, Francisco Doblas-Reyes, Torsten Hoefler, Amy McGovern & Bjorn Stevens

The rapid emergence of deep learning is attracting growing private interest in the traditionally public enterprise of numerical weather and climate prediction. A public–private partnership would be a pioneering step to bridge between physics- and data-based methods, and necessary to effectively address future societal challenges.

Lives and well-beings depend on reliable weather predictions every day. With climate change, the need for better predictions of future weather extremes rises to levels where no investment seems big enough. Weather and climate forecasts are predominantly carried out by nationally governed and globally coordinated public weather services. These forecasts are based on principled equations encoding physical laws, with weather prediction also assimilating hundreds of millions of observations per day. However, the need for weather forecasts and climate projections with increasingly finer granularity and ensemble-based uncertainty quantification causes considerable computational and energy overheads[1]. The costs of such endeavours inevitability hit the public purse.

Private companies first entered this public domain years ago, mostly targeting tailor-made products for niche markets. However, the fast rise of deep learning methods and foundation models, such as Generative Pre-trained Transformer 4 (GPT-4)[2], predominantly driven by private tech companies, are creating unprecedented momentum for not only replacing traditional, physics-based methods, but the entire public service infrastructure with a new type of weather and climate enterprise.

Here, we explore how this momentum shift necessitates a symbiosis between public and private efforts, so that the benefits of deep learning and physics-based models can be exploited to their full potential while preserving the credibility of their products through weather and climate expert input and agreed protocols and quality standards.

The rise of deep learning

Deep learning is a type of artificial intelligence (AI) that uses neural networks to learn tasks from training data. For weather and climate, vast amounts of training data are created from the enormous diversity of model simulations and observations, enabled by the exponential increase in computing and data handling power. For example, the assimilation of weather observations into physics-based forecast models can provide accurate snapshots of the evolving state of the climate system over many decades, documenting global temperature rise and changes in extremes[3]. These high-quality reference datasets[4] make weather and climate prediction primed for deep learning applications and ready for commercial exploitation.

Such datasets are now being used to train deep learning models to make analogue forecasts with impressive skill, promising a factor of 100 × faster time to solution (and even bigger energy-to-solution savings) than would be required by traditional physics-based systems[5–8]. Moreover, by virtue of being constrained by vast amounts of structured observational data, and thereby less influenced by limitations of physics-based models, as well as being amenable to corrections learned from new data sources, these methods should eventually perform better than physics-based models and be more easily tailored to address specific questions (or markets).

To make the training even easier for weather forecasting, it could be possible to bypass physics-based models entirely – the weather on any given day will not abruptly depart from the history of weather on all past days. However, what can be learned depends on the completeness and representativeness of the training data. As such, physics-based models will remain vital, simply because the physics generalizes in ways that are designed to fill in what sparse observations lack to constrain for the fine-grained three-dimensional evolution of weather. Hence, the use of systems built on physical laws will continue to add substantial value to observations.

Deep learning has made lesser inroads for climate prediction, which – following the above line of reasoning – can also be attributed to limitations in the training data. The states of possible future climates cannot be learned in the same way from past states of weather in a slowly evolving climate. Observations of climatically relevant elements are lacking, for instance fractures of ice, the state of the deep ocean or the composition of soils. When it comes to climate, deep learning has the most to offer by incrementally improving physics-based prediction models, and by its ability to extract and interpolate across the state space defined by physics-based model projections. The latter can lead to substantial savings and offer unprecedented, scalable, deep-learning-driven interactivity by the fast reconstruction of physics-based climate trajectory simulations.

Despite this, operational public weather and climate prediction centres have not yet targeted their entire prediction systems for wholesale changes. Rather, centres applied deep learning to well-defined sub-problems of the production workflow (AI inside), with the aim to incrementally improve the quality of their reference datasets[9]. Examples include the emulation of physics-based model components (particularly the costly ones such as calculating the radiation propagation through the atmosphere), deriving bias corrections for forecast models and observations, or the post-processing of model forecasts, data compression and uncertainty quantification.

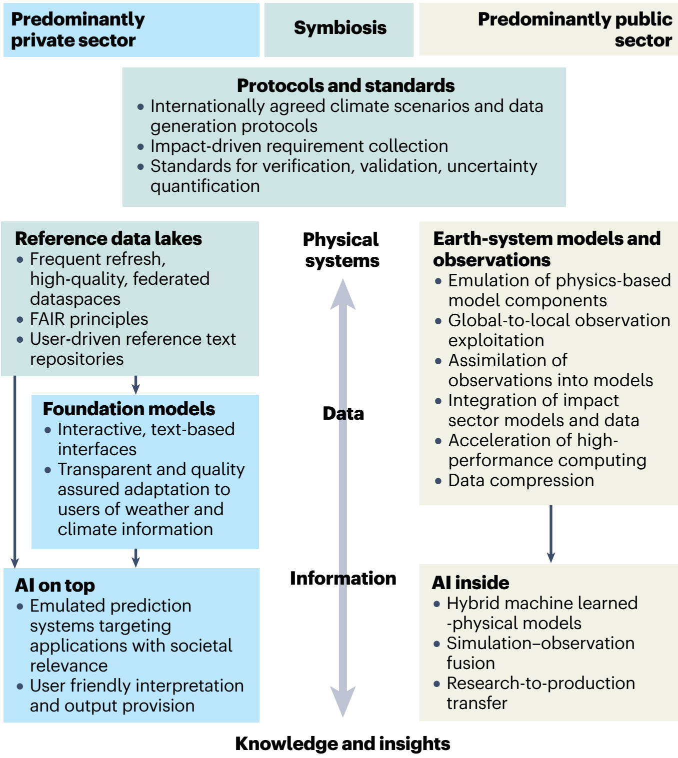

Fig. 1 | Public–private weather and climate information system. Left, the process of developing deep learning and foundation models to utilize reference weather and climate datasets (AI on top). Right, application of deep learning to well-defined sub-problems of the production workflow (AI inside), with the aim to incrementally improve the quality of weather and climate reference datasets. Middle, both sectors should contribute to a transparent process deriving knowledge and insights on climate change and weather extremes from information extracted from data of physical systems. FAIR, Findability, Accessibility, Interoperability, and Reusability.

Foundation models

Foundation models are making another transformative promise – examples including the GPT series or diffusion-based models such as DALL·E2. These models train on vast amounts of unlabelled data using 'predict removed parts' tasks to capture the distribution of the data itself and learn to fill in missing pieces or extend prompts. These learning methodologies closely align with the training of data-driven weather forecasting models, and so domain specific foundation models are expected to develop quickly, particularly driven by private companies.

After identifying basic patterns in the data, foundation models can be fine-tuned to specific tasks. Training for complex tasks, such as conversational agents, can be enabled using Reinforcement Learning from Human Feedback (RLHF), as pioneered in InstructGPT. Developments in the ability of foundation models to learn from human feedback in forecast production[10] means that models could be developed to ingest weather and climate data, which could be fine-tuned by domain experts through RLHF, and also interpret and communicate weather and climate information (AI on top).

The use of foundation models in forecasting would go well beyond physics-based model emulation and could help scale the delivery of this information across sectors. However, the enormity of the global public weather and climate data resource that foundation models exploit emphasizes the importance of creating a public space in this emerging digital universe that helps to control the quality of training and feedback generation – and also sustain scientific credibility.

Outlook

There is an urgent need to rethink the economies of weather and climate prediction given the enormous investments behind efforts to utilize mostly public data, fuelled by a strong push towards deep learning methodologies and powered by effectively unlimited computational and data analytical capacities in the private sector. Foundation models add a new dimension here as they offer unprecedented interpretation and communication capabilities to users.

Exploring the frontiers of deep learning and foundation models thus stands to benefit from an environment where existing public entities are enabled to fully embrace new technologies and quickly gather expertise in domain-specific deep learning – supplemented rather than supplanted by the unique skillsets and infrastructures of 'Big Tech' companies. Public weather and climate services need to continue to drive model development based on physics first principles, fully exploit the information content of observations, and provide public access to the information that can be derived from it.

We suggest that private entities should mostly invest in applying deep learning methods and foundation models to reference datasets, which are generated through transparent processes and that follow agreed standards. Companies would benefit from the generation of deep learning datasets using reference physics, the validation of outputs, and the uncertainty quantification provided by public players, and both public and private sectors should investigate the most innovative solutions for this (Fig. 1).

This symbiosis would help to best tap the full potential of the private sector's innovative spirit and economic power, while ensuring that sufficient investments and input are still made into the underlying public resources. An important ingredient to this symbiosis is the democratization of software and data to enable improved prediction capabilities for all – in particular for the most vulnerable to climate change and weather extremes.

Peter Bauer[1] ✉, Peter Dueben[1], Matthew Chantry[1], Francisco Doblas-Reyes[2,3], Torsten Hoefler[4], Amy McGovern[5,6] & Bjorn Stevens[7]

[1]European Centre for Medium-Range Weather Forecasts (ECMWF), Bonn, Germany. [2]Institució Catalana de Recerca I Estudis Avançats (ICREA), Barcelona, Spain. [3]Earth Sciences Department, Barcelona Supercomputing Center (BSC), Barcelona, Spain. [4]Computer Science Department, Swiss Federal Institute of Technology (ETH), Zürich, Switzerland. [5]School of Computer Science, University of Oklahoma, Norman, OK, USA. [6]National Science Foundation AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES), Norman, OK, USA. [7]Max-Planck-Institute for Meteorology (MPIM), Hamburg, Germany.

✉e-mail: peter.bauer@ecmwf.int

Published online: 1 August 2023

References

1. Bauer, P. et al. The digital revolution of Earth-system science. *Nat. Comput. Sci.* **1**, 104–113 (2021).
2. Brown, T. et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **33**, 1877–1901 (2020).
3. Buontempo, C. et al. The Copernicus Climate Change Service: Climate science in action. *Bull. Amer. Meteor. Soc.* **103**, E2669–E2687 (2022).
4. Dueben, P. D. et al. Challenges and benchmark datasets for machine learning in the atmospheric sciences: Definition, status, and outlook. *Artif. Intell. Earth Syst.* **1**, e210002 (2022).
5. Bi, K. et al. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* https://doi.org/10.1038/s41586-023-06185-3 (2023).
6. Lam, R. et al. GraphCast: Learning skillful medium-range global weather forecasting. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2212.12794 (2022).
7. Pathak, J. et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2202.11214 (2022).
8. Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J. K. & Grover, A. ClimaX: A foundation model for weather and climate. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2301.10343 (2023).

9. Chase, R. J., Harrison, D. R., Burke, A., Lackmann, G. M. & McGovern, A. A machine learning tutorial for operational meteorology. Part I: Traditional machine learning. *Wea. Forecasting* **37**, 1509–1529 (2022).
10. Schick, T. et al. Toolformer: Language models can teach themselves to use tools. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2302.04761 (2023).

Author contributions

P.B. conceived the original concept of the paper, and all authors contributed to the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.