

Digital twins of Earth and the computing challenge of human interaction

Peter Bauer, Torsten Hoefler, Bjorn Stevens & Wilco Hazeleger

 Check for updates

Digital twins of Earth have the capability to offer versatile access to detailed information on our changing world, helping societies to adapt to climate change and to manage the effects of local impacts, globally. Nevertheless, human interaction with digital twins requires advances in computational science, particularly where complex geophysical data is turned into information to support decision making.

Digital twin technology, first conceived and highly evolved for engineering, has expanded its reach into other fields. One of these fields is Earth system science. Connecting the physical Earth system with the adaptation of society to climate change across health, water, food, and energy sectors requires a highly flexible information system, which is where digital twins can shine. However, the complexity of the Earth system, its human component, and the enormous variety of questions that it raises poses new challenges and makes their implementation exceptionally ambitious. While the extreme-scale computing and data analysis aspects of such a twin are becoming understood, the idea of enabling flexible human interaction with a digital twin of Earth is novel and, as we believe, essential.

We consider deep learning methods, and particularly large pre-trained data-driven models (sometimes called ‘foundation models’¹), to be a necessary technology for digital twins of Earth. This will not only require the creation of fast-turnaround Earth system simulators, but also an ability to manage the vast data outputs and to create Earth system and societal impact-specific knowledge databases. Here, ‘instruction models’ (also known as chat bots) can become knowledge interpreters for users from a wide range of backgrounds wanting to interact with Earth data in various ways, by creating generic exploration tools for public users, enabling scientific discovery for experts, and supporting decision making for climate adaptation.

Digital twins of Earth

The key benefits of digital twins of Earth arise from producing high-quality forecasts, reanalysis, and Earth system change projections that can be accessed by a highly interactive system in which users can explore their own scenarios in support of decision making². For this, digital twins must be able to reflect natural and human-made changes in the real world. Some target applications would be to redesign infrastructure or projection to explore the efficacy of the changes in cases where global change creates severe local impacts on the scale of these infrastructures. Others could be to explore how local interventions, for instance related to land and water management, might induce larger

scale hydrological changes³. Yet another application could be to explore the efficacy and side effects from changes meant to influence Earth’s energy budget and hence global warming⁴. Digital twins will be most useful if they can accurately simulate the effects of such changes and thus aid decision making by helping users to explore and assess the impact of proposed actions.

For dealing with present and near-future (days-to-months ahead) questions, digital twins of Earth would help to safely manage the operation of existing infrastructures in health, food, water, energy and other societal sectors in response to present and emerging environmental challenges. For the adaptation to the effects of future climate change (a few decades ahead), digital twins would allow testing new and sustainable environment management solutions and help design next-generation infrastructures. On all-time scales, digital twins would therefore combine physics-based models of the environment with socio-economical and socio-ecological impact models, where the management of impacts can also drive the set-up of the physics-based models if the physical state of the system is affected⁵. It goes without saying that reliable knowledge of how the system will evolve under a given scenario, based on the best available scientific knowledge and methods, is key.

As different applications will need different intervention and expertise levels, computing and data footprints between digital twin of Earth instances can vary substantially. An effective digital twin will be scalable and can also be a system of connected twins, to be managed through intelligent workflows and resource managers to achieve the best trade-off between information quality, timeliness of delivery, and user-developer interactivity.

At the upper end, digital twins of Earth become exascale computing and exabyte data handling problems (meaning, 10^{18} floating point calculations per second and 100s of 10^{15} bytes output per simulation). Part of the computational complexity can be addressed by fundamental rewrites of the software environments for numerical modeling and code adaptation to the latest processor and system technologies⁶. The societal components of the Earth system are less constrained by supercomputing and more by the lack of our ability to understand and generalize social systems⁷. While our digital twins aim to represent the most accurate, first-principles-based digital representation of Earth, real-time user interaction and fast updates will require new methods that create effective shortcuts. In our view, only deep learning can create these efficiency gains for accelerating scientific computing, intersecting between physical and social sciences, and adding interactivity.

Large pre-trained physics-impact models

The computationally heaviest task in digital twins of Earth is the monitoring of past and present change and the prediction and projection of the physical state of climate into the future with physics first-principle-based models. Today’s models already use data assimilation techniques ingesting nearly the entire publicly accessible Earth observation record.

These techniques require ensembles of a few to a few tens of simulations with perturbed initial conditions (and possibly model parameters) to estimate monitoring and prediction uncertainties. Based on recent estimates, the associated need for supercomputing is about 20,000 GPUs, powered by 20 MW to generate sufficient computational throughput⁶. The system should have the flexibility to use its resources to provide fewer simulations at higher spatial resolutions, and shorter time-slices, larger ensembles, or longer simulations at more moderate resolutions. To be successful, these tasks need continued investment in traditional high-performance computing for running complex Earth system simulation codes on centralized hardware installations presently built in Europe, the USA and Japan.

To develop prediction systems that operate at a very small fraction of the cost of physics-based systems, substantial investments in machine learning by model surrogation are being made, and have already been successfully demonstrated for weather prediction⁸. These demonstrators require 4–5 decades worth of past simulation-observation-based weather re-analyses for training. Compared to the above exascale capacity, this training only requires moderate allocations of several hundreds of GPUs over weeks. Once trained, the computing capacity required for the inference step is negligible. Hence, the generation of the training data, rather than the training itself, will determine the computational capacity required of the entire system.

For climate prediction and projection, past climate and weather records will not suffice, because past data is sparse and future climate states are expected to be fundamentally different and lead to future weather states and extremes that have never been observed. We therefore see the biggest role of large data-driven pre-trained models in the interpolation of trajectories across climate snapshots produced by numerical physics-based systems, in fine-graining such trajectories in space, and in translating physical change to societal impacts⁷. The expensive physics-based simulation maintains the general (unknown) trends while the data-driven interpolation creates a set of cheaper ensemble statistics that estimates internal variability generated by nonlinearities in the Earth system⁹. For climate change adaptation, the training and inference steps will be similar to the existing weather prediction examples because several decades worth of multiple (ensemble) simulations at very high to moderate spatial resolution will be needed to train such interpolation: if today's data-driven weather models train with a single model based on 50-year data record, we believe that multi-decadal climate trajectory interpolation models can be trained with 50-year predictions produced by less than 10 models or ensemble members.

Large pre-trained instruction models

Thinking beyond primary data generation and about the development of the above hybrid physical equation–data driven system, the interaction of humans with digital twins also needs to employ large data-driven models, but those that are learning from more than equations and numerical methods (see Fig. 1). Here, the term 'data' needs to be extended, as it will also include specialist datasets produced by commercial companies (for instance, sensors around solar, wind and food farms, car sensors, market indicators, and pricing), by public entities (for instance, traffic cameras, urban air quality sensors, and river monitoring gauges) and even individuals (for instance, citizen scientists and climate consultants), but ultimately also the vast resources of the internet.

The power of these extended types of data-driven models, which we call 'large instruction models', lies in their agility to interact with pretty much any user-specific monitoring and prediction request,

as long as the training data contains the task-specific information. Examples of such models are the well-known BERT and GPT series of models for text, as well as DALL-E and Florence for images. Their fast evolution promises vast opportunities in our domain. To be successful, systems will need to be supported by reinforcement learning, which will require the training of individuals with domain expertise. This has the advantage that such knowledge can be scaled by the instruction models globally. Examples are the tailoring of these models to either harvest or air quality predictions, or to queries by scientists about a specific physical process representation.

As shown in Fig. 1, such models will therefore serve two purposes: (i) to make the physical component of the digital twin of Earth more computable and the resulting vast data outputs manageable (top layer), and (ii) to create a diverse knowledge database as the foundation for the interaction with the twin using language-based instruction models (bottom layer). These two layers will facilitate access to information hidden in complex data and implement the human interface. The role of experts and scientists therefore widens because the system scales their knowledge feedback across many more users and applications.

Computing implications

The top layer in Fig. 1 would learn from abstract climate data and predictions and could be fine-tuned to specific prediction tasks. We would choose a configuration where the model is pre-trained once with numerical simulations, in a very expensive campaign, to obtain a general abstraction for climate data, similar to the weather prediction example trained with re-analyses⁸. It would then be fine-tuned cheaply for producing predictions, uncertainty quantification, or future climate statistics.

To illustrate the computing implications of the extended, large pre-trained instruction model, we chose the set-up of ClimaX¹⁰, as its design principles closely reflect those of so-called foundation models in the weather and climate domain.

Input data would be variable fields sourced from sparse sensor data, as well as regional or global weather forecasts or climate simulations. The different physical variables denote the \times modes in the model. One could use a Vision Transformer Architecture (ViT¹¹) to represent geographical regions and modes. The pre-training objective could be a randomized forecast as in ClimaX. ClimaX uses 48 input variables on a 128×256 grid and an inner dimension of 1,024. With 32-bit floating point precision, the resulting tensor size is 6.4 Gibibytes. ClimaX reduces this burden on memory management by merging the variables into a distributed representation of the inner dimension with a total of approximately 50 million parameters, which represents a small model by today's standards.

The upper limit of computing resources would probably be defined by global km-resolution input data. This would approximately translate to a $17,520 \times 36,000$ grid. Furthermore, we would expect that a large inner dimension (larger than that of ClimaX) would vastly enhance model skill. If we chose GPT-3's 12,288 inner dimension, we would require an input tensor of size $48 \times 17,520 \times 36,000 \times 12,288 = 1.5$ Exbibytes and an accordingly large network with nearly 100 layers in our example. This is clearly not feasible, and one would need to tune the physics data model towards smaller configurations. This could be achieved by coupling the data model with explicit numerical physics simulations to take advantage of the deterministic nature of these simulations or by precomputing simulation data and having the impact model query the data. Other options would include standard artificial intelligence model compression methods such as quantization or

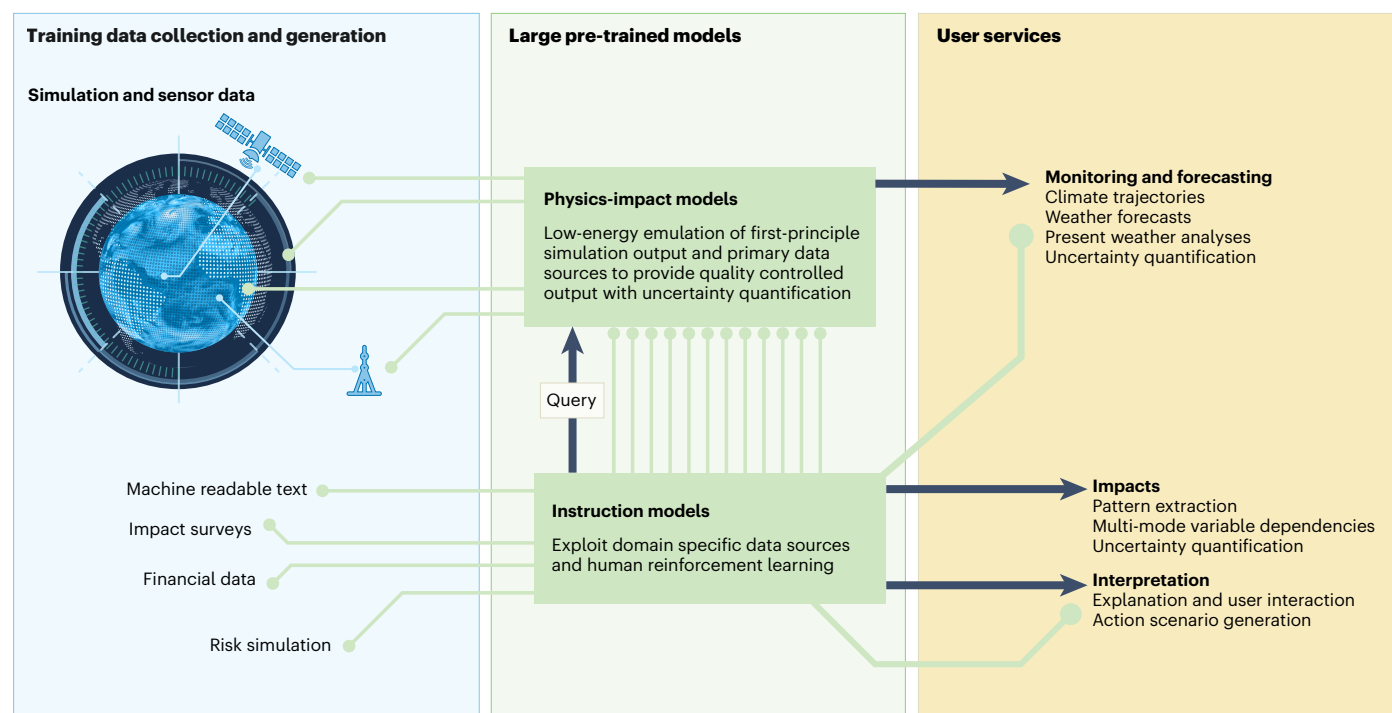


Fig. 1 | Conceptual view of two-layered large-pretrained data-driven modeling system for digital twins of Earth. The figure shows the data production in blue, the artificial intelligence models for physics as well as human instruction in green, and the offered services in yellow. The users query the

models through natural language and other service interfaces. The artificial intelligence models have been trained with the data and may have access to additional data; for instance, the instruction model may query the physics model's data output.

sparsification that may provide 10–100x compression. Lower spatial-resolution input (ensemble) data would also alleviate the input tensor size but would come with other model uncertainties.

The socio-economical and socio-ecological impact, and the instruction model components (the bottom layer in Fig. 1) would query the physics model and data and accept a prompt from a user. The prompt would be written in human language, for instance, “How would Rhine river water flows limit freight traffic for an average year in the 2050s?”; a query that requires insight in global and regional climate change, knowledge on the water management and infrastructures, and knowledge of rules and regulations in at least three countries. The query would address the first part, and perhaps the second part, but likely not the third as it involves law and governance. The model would then interpret the prompt, query the physics model and data, and generate an answer. This represents a daunting task as it requires multiple, interconnected multimodal instruction models. These are only emerging for querying images now¹².

A promising architecture would be to feed a representation of the field (either pre-processed into an embedding by an expert model such as COCO Caption¹³ or directly as in ViT) into a generative pre-trained transformer. OpenAI's GPT-4 has demonstrated promising capabilities in this regard, but the details of the architecture are not public. Early visual instruction models (VIM), similar to MiniGPT-4¹⁴, and LLaVA linked frozen large-language models (LLM) such as LLAMA¹⁵ with image encoders, can achieve visual understanding and question-answering.

One other issue with climate data is the large number of modes (climate variables). This could be addressed by a scheme similar to

ImageBind, in which one mode is used as an anchor to bind the others. ImageBind-LLM¹⁶ has demonstrated promising results for multimodal instruction and conversational agents. Given the dimension of moderate-to-high-resolution ensemble climate data, we expect to require a model with at least several tens of billion parameters, which is similar to what LLaMA is able to manage today.

We could use variable embeddings from the physics model directly (as experts) with an adapter model, or train a separate model as a visual encoder. One could also feed ViT-style tokens directly into the language model prompt. In any case, the model needs to be trained for climate applications, requiring a substantial amount of training data. Fine-tuning to human interaction and instructions requires anywhere between 10,000 and 500,000 interaction examples. We expect our requirements to be at the upper end of this range because climate sciences is a specialized field and thus benefits less from the general internet knowledge base of the LLM component.

While this only illustrates the dimension of the task and identifies possible software solutions, the rapid evolution of this domain promises many opportunities for a fast adaptation for weather and climate applications.

Outlook

The idea of creating a digital twin of Earth with the above-described capabilities would help overcome an imagination deficit that presently impedes effective climate action. Climate data and services have existed for decades, but digital twins enable new ways of creating and interacting with information for scientists, as well as public and private entities tasked with making decisions on matters that affect – and are affected

by – the changing climate. Ideally, the recording of such interactions starts today so that they become learnable data tomorrow.

High-quality science input from simulations and observations sits at the core of digital twins of Earth but must be produced with faster turnaround and with a closer connection to societal impacts and societal impact data, as compared to present practice. Substantial investments in super-computing and emerging digital technologies, but also in science that target deficiencies in the training data, will be necessary to achieve sufficient quality and turnaround when creating physics-based reference and training data.

We believe that deep learning, particularly as an interpreter on top of high-dimensional reference datasets, will be key to realizing our digital-twin vision. This is particularly relevant for adding workability and usability: allowing scientists to perform numerical experiments exploring new knowledge on subsets of such reference datasets, to develop and test methods geared at specific societal impacts, and to work through several adaptation and mitigation strategies. The combination of large pre-trained physics-impact with instruction-type models should bridge across the entire range of digital-twin capabilities and, following our estimates, appear computable. Our community can greatly benefit from the present industrial push for artificial intelligence, but the specific climate application can also create new impetus for artificial intelligence methodological developments applied elsewhere.

It is worth noticing that digital twins of Earth will require substantial computing, thus electrical power for generating training data and, to a lesser extent, for training data-driven models. The electrical power must be generated in the most ecologically sustainable ways. These needs will be offset by the availability and low-power requirements to use the digital twin, so that large pre-trained models will not add to this burden but rather support a more energy-efficient data analysis and feature extraction and create a user interaction platform that would otherwise not exist and be compensated for by probably too many expensive numerical simulations.

The digital technology that creates the computing and data-handling abilities that we need for operating digital twins will only be as powerful as our ability to manage it. This needs a governance framework that is transparent and flexible enough to engage users and become trustable. Important elements are agreed standards for data quality, model quality, and verification, validation, and uncertainty quantification, but also openness of software and data. These can all draw on existing software and data standards and build on existing efforts to create interoperability between heterogeneous infrastructures and disparate data.

The digital twin of Earth concept has been pioneered by the European Destination Earth flagship activity. However, the present

enormous momentum of artificial intelligence should be exploited to make such twins manageable. This thinking sits at the heart of the Earth Virtualization Engine (EVE)¹⁷ initiative that is proposing new ways of creating, managing, and disseminating climate information based on concerted, international investments in this vision.

Peter Bauer ¹✉, **Torsten Hoefler** ²✉, **Bjorn Stevens** ¹ & **Wilco Hazeleger**³

¹Max-Planck-Institute for Meteorology, Hamburg, Germany. ²Computer Science Department, Swiss Federal Institute of Technology, Zürich, Switzerland. ³Faculty of Geosciences, Utrecht University, Utrecht, Netherlands.

✉ e-mail: peter.bauer@mpimet.mpg.de; htor@ethz.ch

Published online: 26 March 2024

References

1. Bommasani, R. et al. Preprint at <https://doi.org/10.48550/arXiv.2108.07258> (2022).
2. Bauer, P., Stevens, B. & Hazeleger, W. *Nat. Clim. Change* **11**, 80–83 (2021).
3. Baldos, U. L. C. et al. *Environ. Res. Lett.* **18**, 053002 (2023).
4. Lawrence, M. G. et al. *Nat. Commun.* **9**, 3734 (2018).
5. Fiedler, T. et al. *Nat. Clim. Change* **11**, 87–94 (2021).
6. Bauer, P. et al. *Nat. Comput. Sci.* **1**, 104–113 (2021).
7. Li, X. et al. *Nat. Rev. Earth Environ.* **4**, 319–332 (2023).
8. Bi, K. et al. *Nature* **619**, 533–538 (2023).
9. Hoefler, T. et al. *Comput. Sci. Eng.* **25**, 50–59 (2023).
10. Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J. K. & Grover, A. Preprint at <https://doi.org/10.48550/arXiv.2301.10343> (2023).
11. Dosovitskiy, A. et al. Preprint at <https://doi.org/10.48550/arXiv.2010.11929> (2021).
12. Ye, Q. et al. Preprint at <https://doi.org/10.48550/arXiv.2304.14178> (2023).
13. Chen, X. et al. Preprint at <https://doi.org/10.48550/arXiv.1504.00325> (2015).
14. Zhu, D., Chen, J., Shen, X., Li, X. & Elhoseiny, M. Preprint at <https://doi.org/10.48550/arXiv.2304.10592> (2023).
15. Gao, P. et al. Preprint at <https://doi.org/10.48550/arXiv.2304.15010> (2023).
16. Han, J. et al. Preprint at <https://doi.org/10.48550/arXiv.2309.03905> (2023).
17. Stevens, B. et al. Preprint at *Earth Syst. Sci. Data* <https://doi.org/10.5194/essd-2023-376> (2023).

Acknowledgements

Yvonne Schrader is thanked for graphical assistance. We thank our editor Fernando Chirigati for improvements of the text. The authors acknowledge the intellectual benefit of exchanges with the broader community of scientists developing EVE.

Author contributions

P.B. proposed the paper and wrote the initial draft, T.H. produced the computing science input, and all authors contributed to the concepts and content of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Peer review information Nature Computational Science thanks Pierre Gentine and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.