Earth Virtualization Engines: A Technical Perspective

Torsten Hoefler, *ETH Zurich, 8092, Zurich, Switzerland*

Bjorn Stevens, *Max Planck Institute for Meteorology, 20146, Hamburg, Germany*

Andreas F. Prein, *National Center for Atmospheric Research, Boulder, CO, 80301, USA*

Johanna Baehr, *Universität Hamburg, 20146, Hamburg, Germany*

Thomas Schulthess, *ETH Zurich and Swiss National Supercomputing Center, 6900, Lugano, Switzerland*

Thomas F. Stocker, *University of Bern, 3012, Bern, Switzerland*

John Taylor, *Commonwealth Scientific Industrial Research Organisation, Canberra, ACT 2601, Australia*

Daniel Klocke, *Max Planck Institute for Meteorology, 20146, Hamburg, Germany*

Pekka Manninen, *CSC-IT Center for Science, 02101, Espoo, Finland*

Piers M. Forster, *University of Leeds, LS2 9JT, Leeds, U.K.*

Tobias Kölling, *Max Planck Institute for Meteorology, 20146, Hamburg, Germany*

Nicolas Gruber, *ETH Zurich, 8092, Zurich, Switzerland*

Hartwig Anzt, *University of Tennessee, Knoxville, TN, 37996, USA*

Claudia Frauen and Florian Ziemen, *German Climate Computing Center, 20146, Hamburg, Germany*

Milan Klöwer, *Massachusetts Institute of Technology, Cambridge, MA, 02139, USA*

Karthik Kashinath, *NVIDIA Corporation, Santa Clara, CA, 95051, USA*

Christoph Schär, *Atmospheric and Climate Science ETH Zurich, 8092, Zurich, Switzerland*

Oliver Fuhrer, *Federal Office of Meteorology and Climatology MeteoSwiss, 8058, Zurich, Switzerland*

Bryan N. Lawrence, *University of Reading, RG6 6UR, Reading, U.K.*

Participants of the Berlin Summit on Earth Virtualization Engines (EVEs) discussed ideas and concepts to improve our ability to cope with climate change. EVEs aim to provide interactive and accessible climate simulations and data for a wide range of users. They combine high-resolution physics-based models with machine learning techniques to improve the fidelity, efficiency, and interpretability of climate projections. At its core, EVEs offer a federated data layer that enables simple and fast access to exabyte-sized climate data through simple interfaces. In this article, we summarize the technical challenges and opportunities for developing EVEs, and argue that they are essential for addressing the consequences of climate change.

We are all witnessing the effects of climate change. Hotter summers, prolonged droughts, massive flooding, and ocean heat waves are examples of extreme weather and climate events that are growing in frequency and intensity. Many agree that addressing climate mitigation and adaptation is the biggest problem humanity faces today. *A large group of scientists and practitioners from different climate-related domains, including some computer scientists, got together for a week in Berlin this*

1521-9615 © 2023 IEEE
Digital Object Identifier 10.1109/MCSE.2023.3311148
Date of current version 25 October 2023.

FROM THE EDITOR-IN-CHIEF

I'm thrilled to bring this timely and important article on Earth virtualization engines (EVEs) to our readers. Climate change is among the most critical issues facing humanity today. Ambitious, cross-disciplinary thinking is needed to tackle it. The EVE vision outlined here—a federated global data infrastructure for open access to high-resolution climate simulations—represents that kind of bold, creative approach. Realizing it poses enormous technical hurdles that also present opportunities for innovations in climate modeling, high-performance computing, machine learning, and open data.

I thank the authors for advancing this vital technical discussion. Their synthesis of ideas from the EVE summit inspires me to think bigger about how computing can tackle society's greatest problem. I hope it galvanizes you as well to bring your expertise to bear on the climate challenge. *Computing in Science & Engineering* is proud to further this important conversation. Let's work to make ambitious visions like EVEs a reality.

—Lorena A. Barba

July to discuss the concept of "Earth virtualization engines" (EVEs). The summit kicked off with this question: "*If climate change is the most critical problem today, why are we not using the largest computers to help solve it?*"

The question proved provocative in two ways. One, many people think we are using them already, although we are not, while others fear that a focus on technologies masks the many other—mostly human and social—dimensions of the problem. Through intense discussions over multiple days, the participants were able to reconcile these different viewpoints in their summary statement. Here, a subset of the participants outline their common understanding of the technical landscape and constraints and a possible technical realization. *This article distills many productive conversations into a set of observations and ideas to form a basis to guide future investments and more detailed investigations*. This is, by no means, meant as a final design; it should be seen as an initial technical contribution to the journey that will enable humanity to assess the detailed risks of climate change from the local to the global scale.

Climate models are currently our central source of information on how climate will continue to change in the future. Thus, any improvement we make to their fidelity, combined with improved theoretical understanding, is likely to directly benefit the quality of climate prediction. Today's exascale computing will enable global climate simulations with a grid resolution approaching 1 km within the next few years. This resolution allows us to explicitly resolve a number of critical physical processes, such as convection in the atmosphere, which are not captured correctly in models typically employed today. Some of the largest improvements when using high-resolution models are seen in the simulation of key variables, such as precipitation, or in capturing orographic effects.[1] However, such kilometer-scale models produce tens of exabytes of data to analyze for diverse uses. Additionally, these *climate models are computationally expensive as well as hard to optimize—let alone use and link to climate impact studies*. This greatly limits the number of groups worldwide that can operate these models, contribute to their development, and benefit from the data they generate.

The Berlin EVE summit recognized that, to cope with the consequences of climate change, many more users have to have equal access to such climate data and the capability to extract information relevant to them. To this end, one of the most critical elements of EVE consists of an *interactive data access layer that allows simple navigation, extraction, and application of the climate simulation data.* In a simple form, we imagine an interface such as Google's Earth Engine or Microsoft's Planetary Computer's "Explore" interface (see Figure 1)—albeit with much better climate predictions.

Behind the scenes, the data of these predictions will be served by a distributed system federating multiple centers running kilometer-scale models as data generators and distilled into actionable information by a layer of custom-built climate tools and services. Such an interactive data layer would also empower local communities—agronomists in India and water managers in Peru—to link their tools to the data. Furthermore, accessing details of the dataset, climate scientists can better understand and reason about simulated changes. Deploying and operating such an interactive system is a grand infrastructure and engineering

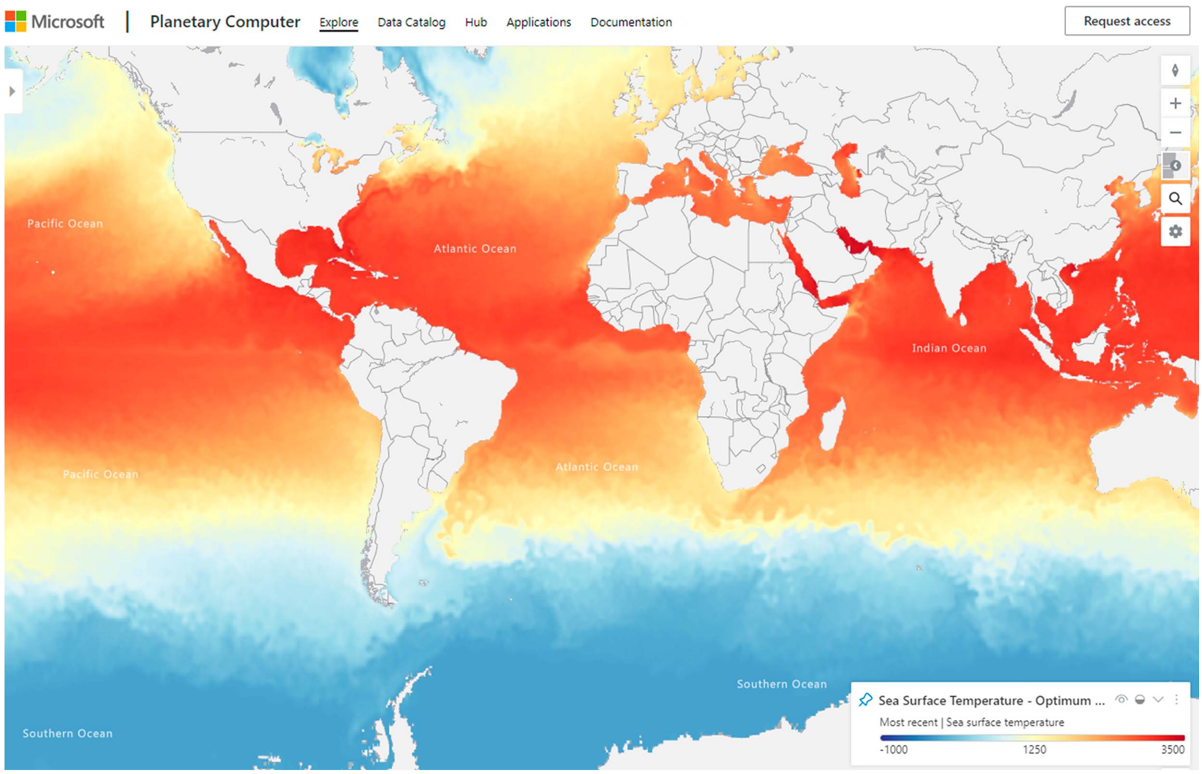

FIGURE 1. Microsoft Planetary Computer Explorer: sea surface temperature. (Source: planetarycomputer.microsoft.com; used with permission.)

challenge. Given its intended use, care must be exercised to ensure that its use is not restricted to industry (e.g., cloud providers, such as Microsoft) or large-scale research laboratories at the forefront of climate research endeavors (e.g., Max Planck or the U.S. Department of Energy).

Machine learning (ML) techniques, which boosted AI with large language models, and related techniques are invigorating climate sciences, as they are many other fields. Summit participants recognized the need to combine physics-based simulations with ML techniques to accelerate and improve predictions, and the scope of data-driven methods was widely discussed. Crucially, ML/AI was identified as essential to opening new opportunities for making the data accessible by *extracting information that can empower users to act.*

WHAT IS CLIMATE MODELING?

Climate is the statistics of weather typically taken over a 30-year period. Weather is what we observe today and what we can predict in the next few days with relatively small uncertainty given accurate initial conditions. However, processes in the atmosphere and ocean are chaotic, which limits our ability to predict the weather beyond two to three weeks.[2] Thus, climate modelers do not aim to predict the weather on a specific day in the distant future but, rather, want to understand how the statistics of weather are changing given alterations in greenhouse gas forcings, aerosols, or other Earth system components. Put differently: what type of weather might we expect in a warmer world?

Predicting the coming weekend's weather can be seen as using physics to extrapolate from today's weather. However, predicting the climate in 30 years, with a potentially doubled CO_2 concentration, requires improved understanding of climate processes. The probability of different weather states may change gradually, but tipping points would accelerate changes, with locally and globally potentially devastating consequences. Understanding; simulating; and, especially, predicting tipping points is very challenging, and Monte Carlo methods are essential to estimate associated uncertainties. Because it is impossible to know exactly what greenhouse gases will be emitted in the future and how humans might respond to the impacts of climate change, simulations span a range of different scenarios.

Another difference from a scientific perspective is how we develop and test tools for climate modeling. We can assess the quality of a weather prediction after some days by comparing it against reality. However, we only have one realization of past climate change that models are typically tuned to capture, and we cannot wait 30 years to understand if our projections were correct. Validating climate projections is, thus, especially tricky, if not impossible, making proper verification of model implementation essential. Thus, it is extremely important to develop our theoretical understanding and, at the same time, tightly control statistics and uncertainty!

HOW DO WE MODEL THE CLIMATE?

Climate simulations are weather predictions run over several decades to centuries but add physical processes relevant for climate projections to explicitly simulate changing CO_2 concentrations. The same governing equations are used for atmospheric processes, and one can share large parts of the code between weather predictions and climate simulations. The principal difference is that coupled atmosphere–ocean simulations are necessary to simulate climate, and full "Earth system models" also include land, ice, and biosphere models. Most of the computational effort is spent in the atmosphere, although this depends on the complexity of other processes included, such as ice sheet and vegetation components.

Many weather forecasting centers participated in the EVE summit, e.g., the European Centre for Medium-Range Weather Forecasts (ECMWF), which offers forecasts as a service. It does this by constantly integrating (or "assimilating"[3]) measurement data from satellites, airplanes, or weather stations into a running simulation as an initial condition for a prediction. ECMWF also uses its data assimilation and modeling capabilities to generate reanalysis products that provide best estimates of the historic weather for the past decades.[4]

IT'S ALL ABOUT THE SCALES!

Climate simulations discretize Earth's atmosphere, ocean, and land surface into a finite set of roughly equally sized tiles, or voxels, which then determine the simulation's resolution. Each grid box represents an average of the physical or chemical properties within. Today's grid resolutions are about 100 km—i.e., we average an area of approximately 10,000 km^2 (or about a quarter of Switzerland) into a single number. This makes it impossible to represent processes that are locally relevant, e.g., because they impact parts of cities. Weather modelers have long ago learned the value of simulating at finer resolution, which is, for instance, why the Swiss National Weather Service is developing weather forecast models with a resolution of kilometers. Coarser grid resolutions cause significant uncertainties in climate simulations. Even global metrics, such as the equilibrium climate sensitivity (ECS) show high uncertainties at current resolutions. ECS assesses the average increase in Earth's surface temperature, assuming the CO_2 in the atmosphere doubled, which ranges from 2 °C to 5 °C, where the uncertainties are mainly due to the representation of cloud changes.[5]

The grid resolution, indeed, has an even bigger impact: physical processes at scales smaller than the grid resolution are not captured by the physical simulation model and need to be represented empirically (using so-called parameterizations). Such parameterizations introduce significant inaccuracies. The accuracy of weather models can be increased significantly by reducing the grid spacing to a few kilometers.[6] Specifically, at such resolutions, the dynamics in deep convective processes, such as large tropical thunderstorms, are simulated explicitly, leading to much higher accuracy in cloud feedbacks.[7] A 100-m grid would further improve the representation of shallow convective clouds, another significant step toward higher fidelity modeling based on first principles, and one particularly relevant for quantities like ECS. Thus, increasing the resolution of climate models is also expected to increase the global prediction quality significantly.

High-resolution climate models are likely our best path forward to improve predictions of the future climate. However, a full Earth simulation at kilometer resolution would require about 510 million tiles for each vertical level and second-scale time stepping over decades of simulated time, making simulations exorbitantly expensive.

SIMULATION COMPUTATIONAL REQUIREMENTS

Running and calibrating high-resolution climate simulations requires the world's best experts working with the world's best research infrastructures. At the same time, much of the insight in climate research is generated in thousands of small-scale research labs by students and research staff who often do not have easy direct access to the simulation setup and data output. Connecting the simulations to the users through frictionless data access interfaces would empower a wide group of researchers and practitioners, help develop and engage talent globally, and fuel new breakthrough ideas.

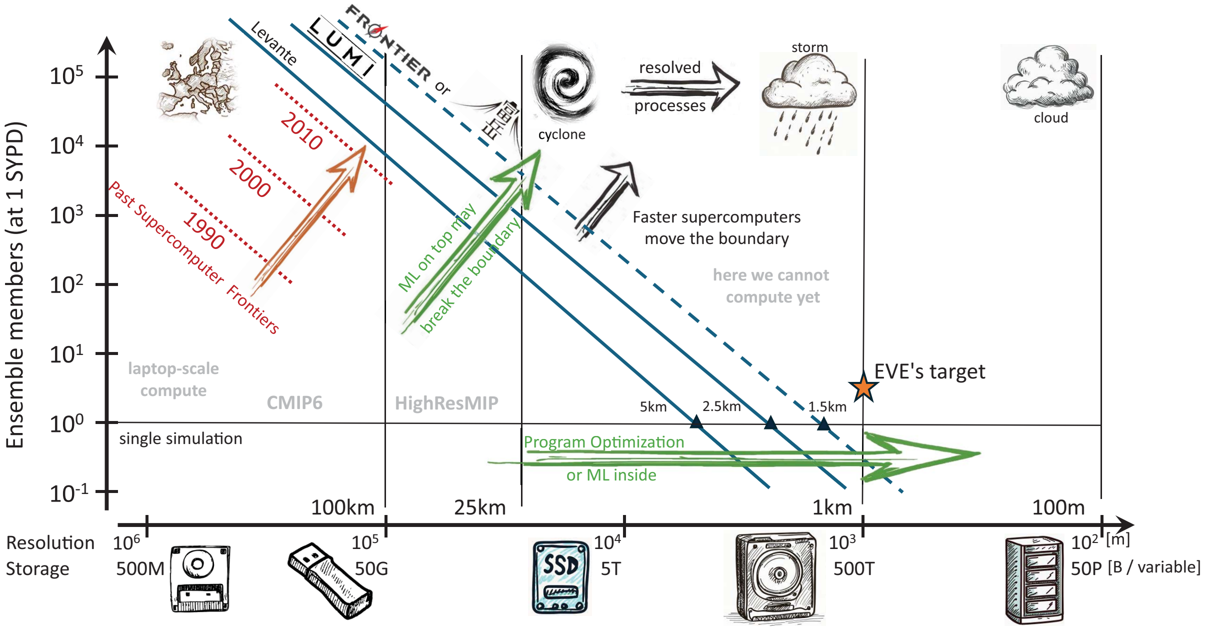

FIGURE 2. Resolution versus number of ensemble members and the computational boundaries to be pushed with EVEs. The x-axis shows increasing resolution, and the number of ensemble members is shown on the y-axis. The x-axis also shows the approximate simulation output data size for one variable over 30 years. A central speed metric is SYPD—it is limited by how simulations perform on today's computers. The illustration shows the limits of existing climate simulations on several machines as blue lines: Icosahedral Nonhydrostatic Weather and Climate Model running on Levante at the German Climate Computing Center, LUMI, and projected on Fugaku and Frontier, which we estimate to perform within 15% of each other for this workload based on their High Performance Conjugate Gradient performance. In red, we outline the historic frontiers based on the available compute capability in the past. We also outline some opportunities for "ML inside" and "ML on top" in green.[14] The typical model resolution today is 100 km; higher resolution models use 25–50 km.[15] EVE: Earth virtualization engine; ML: machine learning; SYPD: simulated years per day.

Furthermore, the relation between increasing the number of ensemble members to capture the probability distribution better and increased resolution to improve the accuracy of each "trajectory" remains complex. Figure 2 shows a sketch of this relationship and several axes of potential improvement. We note that the performance measured in simulated years per day scales linearly with the number of ensemble members but worse with the resolution. Furthermore, the amount of simulation data produced scales rapidly with the resolution. Code optimization or computing speed improvements of the machine will move us to the right in the plot, toward EVE's target. ML "inside" the simulation to replace costly model components offers a potential joker and can lead to exceeding the goal. ML can also combine the benefits of ensembles and resolution in postprocessing "on top" of the simulation to get a better approximation of the distribution.[8]

Based on a rough extrapolation of benchmarking data,[9] we expect that top-class supercomputers, such as Frontier or, soon, Alps, should be able to execute optimized decadal climate simulations at 1-km resolution if the machine was exclusively dedicated to the task for multiple weeks. However, the produced data volume is expected to be many tens of exabytes.[9]

IT'S REALLY ALL ABOUT (ACCESSING AND INTERACTING WITH) THE DATA!

From a technical perspective, a climate simulation produces an output array of values at each grid point at specific output intervals (e.g., 15 min). The potentially many ensemble members sample the probability distribution of each of those values. Therefore, a simulation's output can be identified by a multidimensional tensor: the horizontal i, j (surface) dimensions of the

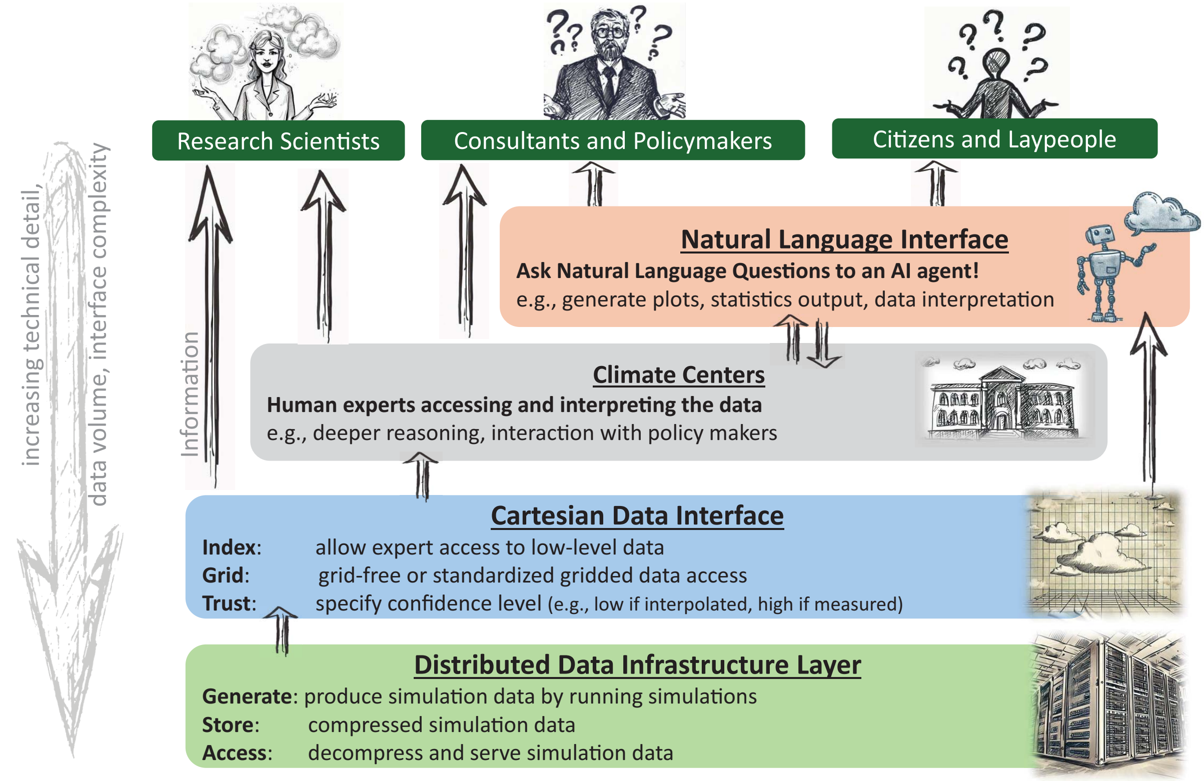

FIGURE 3. Information flow in a potential EVE data layer design.

grid; the vertical k (atmospheric or ocean) dimension; the time dimension t; and the ensemble dimension e for each variable of interest (e.g., temperature).

A single variable, for example, the surface temperature, stored as 8-b value sampling in 15-min intervals for a typical 30-year simulation requires 0.5 PiB of storage. However, the temperature at different height levels controls cloud formation, such that one would need to store more than 100 times this value in the vertical dimension.[9] Because the climate is described by many variables (tens to hundreds), and tens of simulations are required to sample internal variability, naive output strategies can easily produce many exabytes to characterize a single climate change scenario.

Given the exabyte storage requirements, EVE must offer a storage back end that supports highly optimized, domain-specific compression.[10,11] Furthermore, EVE's access protocols must be extremely lightweight and efficient to enable the fastest low-overhead access to the data. Most importantly, such a planetary, federated data-serving ecosystem for climate data that is globally accessible simplifies the rapid development of ML applications on top and avoids data copies. In fact, simple access through a global index and compatible access protocols will foster collaboration, reproducibility, and progress.

We envision a simple interface that can be accessed by users with very different levels of sophistication (Figure 3). EVE's data layer has different access roles: scientists, consultants and policymakers, and citizens and laypeople—each of which have different requirements. Experts want access to all data fields and can deal with vast unfiltered gridded variable data; consultants want interpreted access to the simulation data; and citizens want analyses of interest in a digestible format, such as maps of the risk of flooding or wildfires.

The raw data can be accessed through the Cartesian Data Interface as a sparse multidimensional tensor. The dimensions include the spatial dimensions i, j, and k of the grid and additional ones to index the variable of interest, generating the model and model configuration. The tensor is sparse to also support observation data and fill in nonexisting information. The latter could be offered using automated (e.g., ML-based) interpolation of nonexisting elements and return an uncertainty for each queried point. This results in a query interface similar to getPoint (i, j, k, t, e, var, model, setup, and so on) that

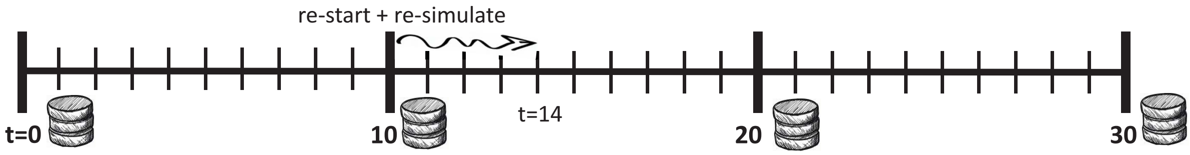

FIGURE 4. SimFS resimulation. We show 30 output steps in 15-min intervals, each consisting of many more simulation time steps (not shown). During the simulation, data are only written at output steps divisible by 10, thus providing a 10× compression ratio. When, for example, step 14 is required, the simulation would be restarted at step 10 and run for four output steps forward.[12]

returns a point and its certainty. Similarly, region queries return subtensors or neighborhoods. It could be served through RESTful application programming interfaces (APIs), like ECMWF's polytope server, or a faster low-level cluster API using remote direct memory access, for example. This could be invisible to users if EVE offers language bindings. A discovery interface would allow the ability to enumerate dimensions such as all available simulation setups. Unifying this storage of climate information would require agreement on the same grid, e.g., HEALPix.

Developing the data layer for EVE will be challenging but with exciting opportunities for innovation, especially for the application of ML. For example, instead of compressing the data, checkpoints at specific time intervals as proposed by SimFS[12] could be used for resimulations (Figure 4) or ML-generated trajectories that faithfully reproduce the distribution of simulations between checkpoints (Figure 5). Alternatively, innovative ML-based compression techniques could overfit model data to reproduce a tensor of data from its coordinates.[10] Furthermore, users could preregister analyses for a large simulation run that will be executed on the fly, similar to beamlines in CERN.[6]

INITIAL IDEAS FOR USING ML TECHNIQUES

Three main avenues for using ML techniques in the context of EVE are possible: 1) improving the simulation, 2) postprocessing and storing the data, and 3) interpreting the data.

Ideas for improving the simulation itself abound. One approach is to replace parameterized parts of the simulation, such as radiation,[13] with faster and potentially more accurate ML-based models. Another, more extreme, idea is to rely purely on data-driven forecasting and use ML-based weather models, such as Pangu-Weather or FourCastNet. However, those may diverge quickly from reality if they are not constrained by physical laws, especially when used to extend into future climate states. An intermediate approach would be to use a SimFS-style checkpointing method running a simulation forward and storing checkpoints every k steps. Then, one would train a "tethered" model to converge to each of those checkpoints like a tethered boat or cable ferry. At the same time, the model could capture the statistics of multiple trajectories between the checkpoints (Figure 5). Of course, the tethering points themselves should follow ensemble members from the overall distribution. The benefits are twofold: the trained ML-tethering model generates trajectories 100–1000 times faster than a traditional climate simulation while simultaneously offering 100–1000-times compression on data storage requirements.

A second avenue for ML use is postprocessing to optimize data storage. One way to compress data would be to explicitly model the uncertainty in the ensemble dimension using simple statistics or advanced

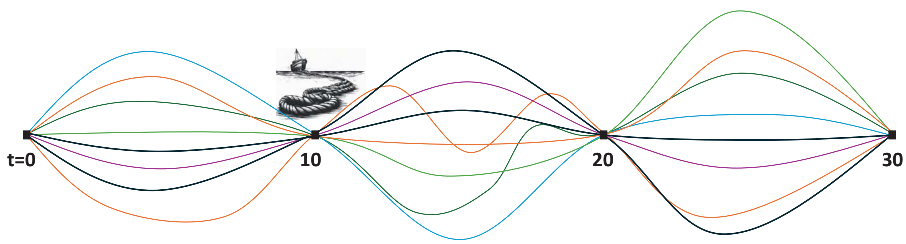

FIGURE 5. Eight colored trajectories are tethered at four points at 10-output-step intervals. This way, the overall simulation's drift is controlled while representing the statistics. ML methods, such as generative AI, could increase the number of steps between intervals to 100–1000.

ML models.[8] Another opportunity could be to use ML models to compress the fields in the tensor directly by overfitting a model to reconstruct a block of the tensor data as exactly as possible. Earlier work has demonstrated compression rates of 1000 times and more using this technique.[10]

The most important and probably biggest avenue for using ML techniques is interpreting the data, i.e., extracting actionable information[14] At the highest level, an interactive "interpretative agent" large language model would be able to answer questions about climate scenarios in natural language, potentially supported by output plots and data series. Figure 3 shows this as the natural language interface, which could build on other ML models that extract information in a hierarchical way at lower levels. An exciting opportunity here is that ML not only helps achieve interactivity but also massively scales the extraction of actionable information in myriad user-defined, bespoke ways—exactly what is needed to realize EVE's vision. This area is largely unexplored but offers significant opportunities.

For each of those tasks, one could think about training foundation models that represent the latent distribution of climate data. One could store the inherent properties of climate data in an ML-derived latent space that could lead to faster compression using fine-tuning. A second foundation model could be the knowledge base for the interpretative agent language model.

CONCLUSION AND OUTLOOK

EVEs stand as a critical intersection between computer, computational, and climate sciences, necessitating collaboration among these sectors, industry, and academia. The technical challenges EVE poses are enormous—from creating high-resolution simulations to making exabyte-level data accessible for all. However, these hurdles represent remarkable opportunities for innovation and cross-disciplinary collaboration. Together, we can make EVE a reality, not just a vision.

ACKNOWLEDGMENTS

We thank all participants of the EVE summit in Berlin for many engaging discussions! Torsten Hoefler was supported by the ADIA lab. We thank Lorena Barba for her editorial comments improving the quality of the manuscript.

REFERENCES

1. A. F. Prein et al., "A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges," *Rev. Geophys.*, vol. 53, no. 2, pp. 323–361, Apr. 2015, doi: 10.1002/2014RG000475.
2. E. Lorenz, "*Predictability: Does the flap of a butterfly's wing in Brazil set off a tornado in Texas*?" 1972. [Online]. Available: https://mathsciencehistory.com/wp-content/uploads/2020/03/132_kap6_lorenz_artikel_the_butterfly_effect.pdf
3. E. Klinker et al., "The ECMWF operational implementation of four-dimensional variational assimilation. III: Experimental results and diagnostics with operational configuration," *Quart. J. Roy. Meteorol. Soc.*, vol. 126, no. 564, pp. 1191–1215, Apr. 2000, doi: 10.1002/qj.49712656417.
4. H. Hersbach et al., "The ERA5 global reanalysis," *Quart. J. Roy. Meteorol. Soc.*, vol. 146, no. 730, pp. 1999–2049, Jul. 2020, doi: 10.1002/qj.3803.
5. P. Forster et al., "The Earth's energy budget, climate feedbacks, and climate sensitivity," in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte et al., Eds. Cambridge, U.K.: Cambridge Univ. Press, 2021, pp. 923–1054, doi: 10.1017/9781009157896.009.
6. C. Schär et al., "Kilometer-scale climate models: Prospects and challenges," *Bull. Amer. Meteorol. Soc.*, vol. 101, no. 5, pp. E567–E587, May 2020, doi: 10.1175/BAMS-D-18-0167.1.
7. T. Schneider et al., "Climate goals and computing the future of clouds," *Nature Clim. Change*, vol. 7, no. 1, pp. 3–5, Jan. 2017, doi: 10.1038/nclimate3190.
8. P. Grönquist et al., "Deep learning for post-processing ensemble weather forecasts," *Philos. Trans. Roy. Soc. A*, vol. 379, no. 2194, Feb. 2021, Art. no. 20200092, doi: 10.1098/rsta.2020.0092.
9. T. C. Schulthess et al., "Reflecting on the goal and baseline for exascale computing: A roadmap based on weather and climate simulations," *Comput. Sci. Eng.*, vol. 21, no. 1, pp. 30–41, Jan./Feb. 2019, doi: 10.1109/MCSE.2018.2888788.
10. L. Huang and T. Hoefler, "Compressing multidimensional weather and climate data into neural networks," 2022, *arXiv:2210.12538*.
11. M. Klöwer et al., "Compressing atmospheric data into its real information content," *Nature Comput. Sci.*, vol. 1, no. 11, pp. 713–724, Nov. 2021, doi: 10.1038/s43588-021-00156-2.
12. S. Di Girolamo et al., "SimFS: A simulation data virtualizing file system interface," *in Proc. IEEE Int. Parallel Distrib. Process. Symp. (IPDPS)*, 2019, pp. 621–630, doi: 10.1109/IPDPS.2019.00071.
13. F. Chevallier et al., "Use of a neural-network-based long-wave radiative-transfer scheme in the ECMWF

atmospheric model," *Quart. J. Roy. Meteorol. Soc.*, vol. 126, no. 563, pp. 761–776, Dec. 2006, doi: 10.1002/qj.49712656318.
14. P. Bauer et al., "Deep learning and a changing economy in weather and climate prediction," *Nature Rev. Earth Environ.*, vol. 4, no. 8, pp. 1–3, Aug. 2023, doi: 10.1038/s43017-023-00468-z.
15. R. J. Haarsma et al., "High resolution model intercomparison project (HighResMIP v1.0) for CMIP6," *Geoscientific Model Develop.*, vol. 9, no. 11, pp. 4185–4208, Nov. 2016, doi: 10.5194/gmd-9-4185-2016.

TORSTEN HOEFLER is a professor of computer science at ETH Zurich, 8092, Zurich, Switzerland. His research interests include large-scale AI and high-performance computing (HPC) systems as well as applications in the areas of large language models and climate sciences. Hoefler received his Ph.D. degree in computer science from Indiana University. Contact him at htor@ethz.ch.

BJORN STEVENS is the managing director of the Max Planck Institute for Meteorology, 20146, Hamburg, Germany, and a professor at the University of Hamburg. His research interests include the physics of climate. Stevens received his Ph.D. degree in atmospheric science from Colorado State University. Contact him at bjorn.stevens@mpimet.mpg.de.

ANDREAS F. PREIN is a researcher at the National Center for Atmospheric Research, Boulder, CO, 80301, USA. His research interests include high-resolution modeling of mesoscale processes in the climate system, including extreme weather events. Prein received his Ph.D. degree in physics from the University of Graz. Contact him at prein@ucar.edu.

JOHANNA BAEHR is a professor at the Institute of Oceanography at Universität Hamburg, 20146, Hamburg, Germany, where she leads a working group on "Climate Modeling," which focuses on subseasonal to decadal climate prediction. Baehr received her Ph.D. degree in physical oceanography from Universität Hamburg. Contact her at johanna.baehr@uni-hamburg.de.

THOMAS SCHULTHESS is a professor of computational physics at ETH Zurich and the director of the Swiss National Supercomputing Center, 6900, Lugano, Switzerland. Schulthess received his Ph.D. degree in natural science from ETH Zurich. Contact him at schulthess@cscs.ch.

THOMAS F. STOCKER is a professor of climate and environmental physics with the Physics Institute at the University of Bern, 3012, Bern, Switzerland. His research interests include the dynamics of the climate system over the past 1 million years and its future evolution, particularly anthropogenic climate change. Stocker received his Ph.D. degree in natural sciences from ETH Zurich. Contact him at stocker@climate.unibe.ch.

JOHN TAYLOR is a research group leader with Data61, Commonwealth Scientific Industrial Research Organisation, Canberra, ACT 2601, Australia; the chief computational scientist at the Defence Science and Technology Group; and an honorary professor with the College of Engineering and Computer Science, Australian National University. His research interests include AI for science, computational and simulation science, and climate change. Contact him at john.taylor@data61.csiro.au.

DANIEL KLOCKE is the leader of the Computational Infrastructure and Model Development Group at the Max Planck Institute for Meteorology, 20146, Hamburg, Germany. His research interests include the development of the next generation of climate models with kilometer-scale resolution and their application on large HPC systems. Klocke received his Ph.D. degree in meteorology from the University of Hamburg. Contact him at daniel.klocke@mpimet.mpg.de.

PEKKA MANNINEN is the director of science and technology at the Advanced Computing Facility at CSC-IT Center for Science Ltd., 02101, Espoo, Finland, and an adjunct professor at the University of Helsinki. His research interests include supercomputing and supercomputing infrastructures. Manninen received his Ph.D. degree in theoretical physics. Contact him at pekka.manninen@csc.fi.

PIERS M. FORSTER is a professor of physical climate change and director of the Priestley International Centre for Climate at the University of Leeds, LS2 9JT, Leeds, U.K. His research interests include the causes and impacts of climate change. Piers received his Ph.D. degree in meteorology from the University of Reading. Contact him at p.m.forster@leeds.ac.uk.

TOBIAS KÖLLING is a postdoc with the Max Planck Institute for Meteorology, Hamburg, 20146, Germany. His research interests include remote sensing, climate physics, and attached data systems. Kölling received his doctoral degree in natural sciences from Ludwig-Maximilians-Universität in Munich. Contact him at tobias.koelling@mpimet.mpg.de.

NICOLAS GRUBER is a professor of environmental physics at ETH Zurich, 8092, Zurich, Switzerland. His research interests

include Earth's biogeochemical cycles, especially that of carbon, and using models and observations to determine how these cycles interact with the physical climate system. Gruber received his Ph.D. degree in environmental physics from the University of Bern. Contact him at nicolas.gruber@env.ethz.ch.

HARTWIG ANZT is the director of the Innovative Computing Lab and a professor with the Electrical Engineering and Computer Science Department, University of Tennessee, Knoxville, TN, 37996, USA, as well as a senior research scientist at the Steinbuch Centre for Computing, Karlsruhe Institute of Technology. His research interests include numerical methods for next-generation hardware architectures. Anzt received his Ph.D. degree in applied mathematics from the Karlsruhe Institute of Technology. Contact him at hanzt@icl.utk.edu.

CLAUDIA FRAUEN is a research software engineer at the German Climate Computing Center (DKRZ), 20146, Hamburg, Germany. Her research interests include HPC for weather and climate with a focus on strategies for performance portability to enable kilometer-scale climate simulations. Frauen received her Ph.D. degree in climate science from Christian-Albrechts-University Kiel and the GEOMAR Helmholtz Center for Ocean Research Kiel. Contact her at frauen@dkrz.de.

FLORIAN ZIEMEN is a research software engineer at DKRZ, 20146, Hamburg, Germany. His research interests include the analysis and visualization of climate model output on HPC systems. Ziemen received his Ph.D. degree in climate science from the University of Hamburg. Contact him at ziemen@dkrz.de.

MILAN KLÖWER is a postdoctoral associate at the Massachusetts Institute of Technology, Cambridge, MA, 02139, USA. His research interests include climate model development, low-precision computing, and climate data compression. Klöwer received his Ph.D. degree in climate computing from the University of Oxford. Contact him at milank@mit.edu.

KARTHIK KASHINATH is a principal engineer and scientist at NVIDIA Corporation, Santa Clara, CA, 95051, USA, and coleads NVIDIA's Earth-2 Initiative. His research interests include large-scale AI and HPC as well as physics-informed machine learning. Kashinath received his Ph.D. degree in aerospace engineering from the University of Cambridge. Contact him at kkashinath@nvidia.com.

CHRISTOPH SCHÄR is professor at the Institute for Atmospheric and Climate Science at ETH Zurich, 8092, Zurich, Switzerland. His research interests include developing and exploiting high-resolution regional climate models on HPC platforms. Schär received his Ph.D. degree in atmospheric dynamics from ETH Zurich. Contact him at schaer@env.ethz.ch.

OLIVER FUHRER is the head of numerical prediction with MeteoSwiss, 8058, Zurich, Switzerland. His research interests include high-resolution numerical weather prediction and HPC for weather and climate. Fuhrer received his Ph.D. degree in physics from ETH Zurich. Contact him at oliver.fuhrer@meteoswiss.ch.

BRYAN N. LAWRENCE is the University of Reading Professor of Weather and Climate Computing, University of Reading, RG6 6UR, Reading, U.K., and a senior scientist in the U.K. National Centre for Atmospheric Science. His research interests include atmospheric dynamics, climate science, and modeling and data technologies. Lawrence received his Ph.D. degree in atmospheric physics from the University of Canterbury in New Zealand. Contact him at bryan.lawrence@ncas.ac.uk.