

Cinema image-based *in situ* analysis and visualization of MPAS-ocean simulations



Patrick O'Leary^{a,*}, James Ahrens^b, Sébastien Jourdain^a, Scott Wittenburg^a,
David H. Rogers^b, Mark Petersen^b

^a Kitware, Inc. 1800 Old Pecos Trail, Suite G, Santa Fe, 87505 NM, United States

^b Los Alamos National Laboratory, Los Alamos, 87545 NM, United States

ARTICLE INFO

Article history:

Available online 21 October 2015

Keywords:

High performance computing

In situ

Image-based

Analysis and visualization

ABSTRACT

Due to power and I/O constraints associated with extreme scale scientific simulations, *in situ* analysis and visualization will become a critical component to scientific exploration and discovery. Current analysis and visualization options at extreme scale are presented in opposition: write files to disk for interactive, exploratory analysis, or perform *in situ* analysis to save data products about phenomena that a scientist knows about in advance. In this paper, we demonstrate extreme scale visualization of MPAS-Ocean simulations leveraging a third option based on Cinema, which is a novel framework for highly interactive, image-based *in situ* analysis and visualization that promotes exploration.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Extreme scale ($\geq 10^{15}$ FLOPS) parallel computing systems require scientists to change the manner in which they analyze and visualize simulation results. Until recently, data, at significantly smaller scales, have been primarily stored and, typically, moved to other machines for post-processing. Transitioning to next-generation computing systems means that these standard post-processing techniques will not be able to effectively scale due to availability of storage and the relatively slow increase in both storage and network bandwidth.

In situ techniques that analyze and visualize simulation data in simulation memory are at least one promising path forward. In general, *in situ* approaches operate on a predefined set of analyses and visualizations. In order for *in situ* analysis and visualization to meet the needs of scientists at extreme scale, it should: (1) preserve important elements of the simulations, (2) significantly reduce the data needed to preserve these elements, and (3) offer as much flexibility as possible for post-processing exploration.

In this paper, describing the visualization of a large-scale Model for Prediction Across Scales (MPAS) Ocean simulation, we leveraged the Cinema framework [1,2]. Cinema is a novel framework for highly interactive, image-based *in situ* analysis and visualization that promotes interactive exploration. This interactive exploration – so important to scientific discovery – is supported intuitively. Cinema effectively preserves this capability by allowing *in situ* data products to be combined in much the same way they could be in a post-processing tool like ParaView.

Cinema imagery is on the order of 10^6 in size, whereas extreme scale simulation data is on the order of $\geq 10^{15}$ in size. This provides the scientist with 10^9 , or greater, images to facilitate capture, curation, and exploration of important elements (features) in the simulation. Cinema is an *in situ* exploration tool used in a traditional *in situ* operational mode, but it allows us to sample

* Corresponding author. Tel.: +15 183713971.

E-mail address: patrick.oleary@kitware.com (P. O'Leary).

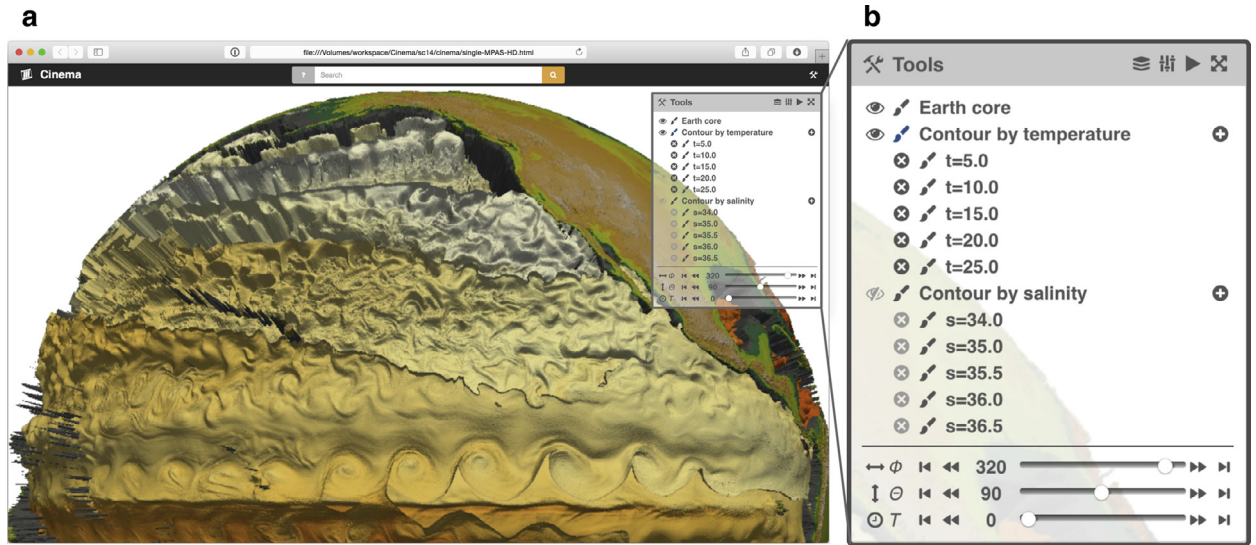


Fig. 1. (a) MPAS-Ocean simulation indicating isosurfaces that represent the locations of water masses within the ocean of temperature colored by salinity. (b) The user interface of the visualization pipeline for analysis and visualization objects (Earth core, temperature contours, and salinity contours) compositing.

the analysis and visualization parameter space, such as camera positions, analysis and visualization objects, time, etc., to produce a set of images stored in a data-intensive database. A positive attribute of Cinema is derived from the fact that the analyses and visualizations are created from full-resolution data with high accuracy.

The visualization shown in Fig. 1(a) (as well as a video available at <https://vimeo.com/127075375>) is a run of MPAS-Ocean. MPAS [3] is a new software framework for the rapid development of climate model components on unstructured grids. MPAS-Ocean provides large-scale parallel scalability, expected in a next-generation code, and support for variable resolution meshes. MPAS is developed cooperatively by Los Alamos National Laboratory and the National Center for Atmospheric Research for the purpose of decadal to century-long climate change research, as well as short-term weather forecasting.

The high-resolution global ocean simulation with realistic topography, uses 1.8 million horizontal 15 km-sized grid cells and 40 vertical levels. The isosurfaces of temperature and salinity indicate the locations of water masses within the ocean. Water masses, with names like North Atlantic Deep Water and Antarctic Bottom Water, occur within specific ranges of temperature and salinity. These visualizations allow oceanographers to view the pathways and extents of these water masses, and compare them to observed climatology. Meandering ocean jets and eddies are visible as perturbations to these analysis and visualization objects, and the ocean currents and eddy activity compare well with observations.

These simulations are typically run on 6000 processors, and achieve two simulated years per wall clock day. At this resolution, file output sizes present difficulties for traditional post-processing analysis and visualization workflows.

2. Related work

Cinema makes a number of contributions, while building on a vast amount of previously disseminated results in interactive exploration databases, creating new visualizations and metadata and content searching.

Interactive exploration database. A large image collection is produced from a structured sampling of time, analysis and visualization objects and camera positions. One option for the management of this large image collection is to compress them into movies [4,5]. Cinema extends this approach by support compositing of images to create new visualizations as well as metadata and content searching. Tikhonova et al. [6–8] represent data as a collection of proxy images. Combining the two approaches would support additional data reduction and exploration possibilities.

Creation of new visualizations. Cinema supports combining images using depth ordering information, while Tikhonova et al. [6] creates new visualizations using interpolation. Cinema's idea for compositing visualization results evolved from the long history of parallel compositing techniques that enable scalable interactive visualization [9].

Metadata and content searching. Recent work by Subrahmaniam [10] identifies issues and future research directions for multimedia databases. Cinema creates imagery *in situ* saving into its database: camera positions, time, details about the analysis and visualization objects, and statistics about the data. Directly storing analysis results and how they are created is important and encompasses our connection to provenance systems [11].

Image content queries support querying about the visual weight of the objects in the generated visualization. There are many approaches to calculating the statistics of the 2D projection of a set of 3D objects [12]. Related to our approach, Jun Tao

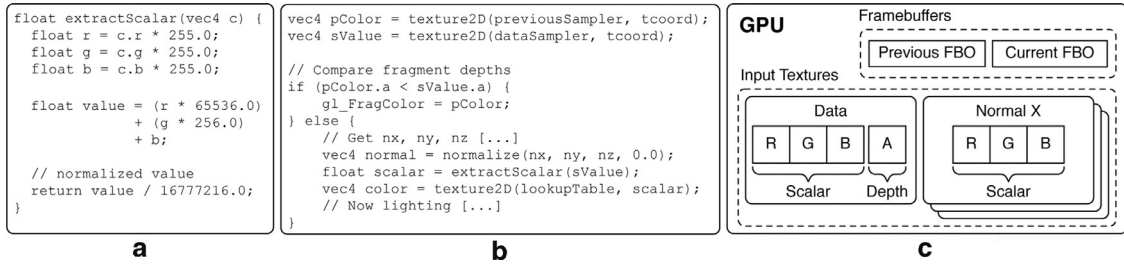


Fig. 2. The interactive exploration, enabled by the Cinema graphical user interface (GUI), pulling resulting Cinema images and analysis products from the interactive exploration database.

et al. [13] computes a collection of streamline images, and applies an image quality metric to select an optimal viewpoint. Our approach extends this work by virtue of being *in situ* and our ability to change our evaluation metric dynamically with a scientist-generated query.

3. Discussion

Interactive exploratory post-processing analysis and visualization will still be essential for scientific discovery at extreme scale. First, debugging new simulators requires this support. Second, collaboration for colleagues without equal computing resources need this support. Finally, transitioning existing simulators demands this support for a path from terascale/petascale ($10^{12}/10^{15}$ FLOPS) to extreme scale.

3.1. From simulation data to image database

Cinema, currently, leverages ParaView [14] to define a desired set of analysis and visualization objects based on a test data set. At this point, scientists can make sophisticated prioritized decisions for the production of Cinema output. They can define how to sample: time, analysis and visualization object parameter space, camera space, and image resolution. The results of these choices are reported by cost metrics associated with the number of images, the image size, the imagery size, and additional computational time.

As simulations progress towards extreme scale, scientists will operate in a constrained storage and bandwidth environment. Hence, sampling of simulation results will require prioritization to fit within a storage/bandwidth budget. Cinema was designed to operate within these constraints, and can assist scientists in making appropriate decisions.

For the *in situ* analysis and visualization, we use Catalyst, an open-source *in situ* data analysis and visualization optimized C++ library, which includes an API for C, FORTRAN, Python and C++. Catalyst is designed to be tightly coupled by directly embedding it into parallel simulation codes.

As the Catalyst enabled simulation runs, the Cinema results are ingested into a database. By this we mean, that the metadata, image provenance (i.e., a searchable description of how the image was created – the simulation, the input deck, and which objects were created), and the image uniform resource locator (URL). Therefore, there is potentially no need to move the potentially large image data around in bulk.

3.2. Creation of new visualizations

Cinema utilizes real-time compositing to create an experience reminiscent of interactively exploring the raw simulation data. For the Web-based Cinema viewer, compositing is enabled through WebGL using ping-pong shaders. For each object (layer) to be composited (see Fig. 2(c)), Cinema loads the data and surface normals into textures. The previous framebuffer object (FBO) is bound to a texture for reading, while the current FBO is bound for writing (rendering). The object is composited by drawing a textured rectangle, and then swapping the current and previous FBOs.

The field data associated with an analysis and visualization object is stored as a 24-bit float in the r, g, and b components of separate Cinema images. The alpha component stores the depth information. In Fig. 2(a), we demonstrate how the scalar value is extracted from the object data texture. Finally, as depicted in the Fig. 2(b) fragment shader, for each fragment the depth is tested, and if compositing is required, the normal and scalar components are extracted and used to compute a normal and color for Blinn–Phong lighting calculations.

It's Cinema's user interface that significantly enhances the exploration experience seemingly beyond those available in traditional post-processing tools. From a scientist's perspective, computationally intensive analysis and visualization algorithms run in constant time using the precomputed Cinema images and analysis products. The interactive exploration database currently supports three elements and two modes of interaction. The three elements of interaction are *time*, *analysis* and *visualization objects*, and *camera*. The modes of interaction are: *animation*; where the interaction is a sequence through time, objects and camera; and *selection*, where the scientist selects or queries time, objects, and camera.

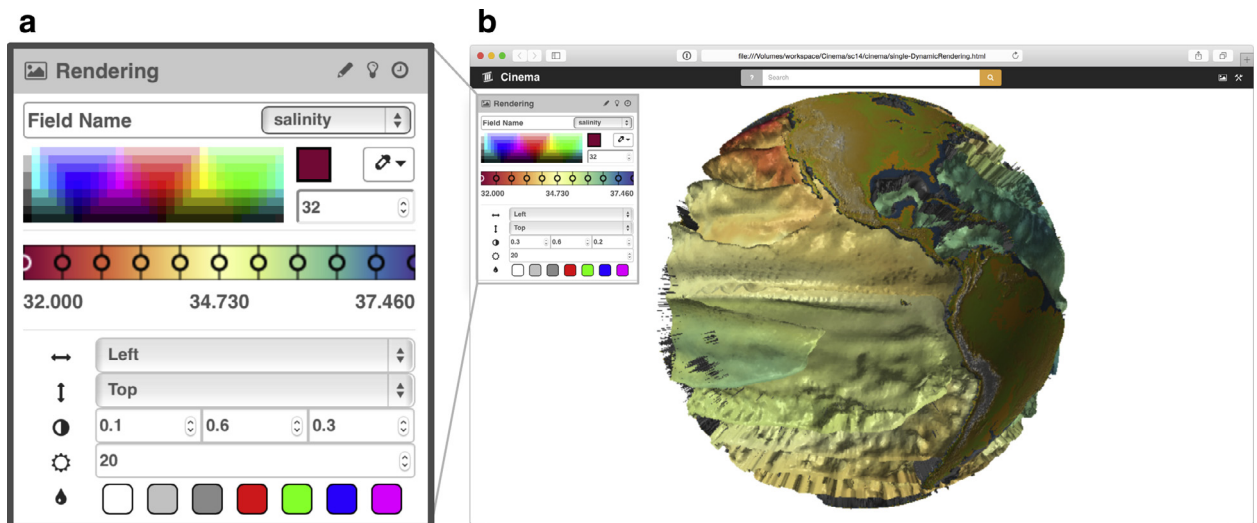


Fig. 3. (a) The dynamic rendering widget offers the capability to edit color lookup tables, set the light direction and color, and adjust material properties. (b) The fully interactive exploration, enabled by the Cinema GUI, pulling resulting Cinema images and analysis products from the interactive exploration database.

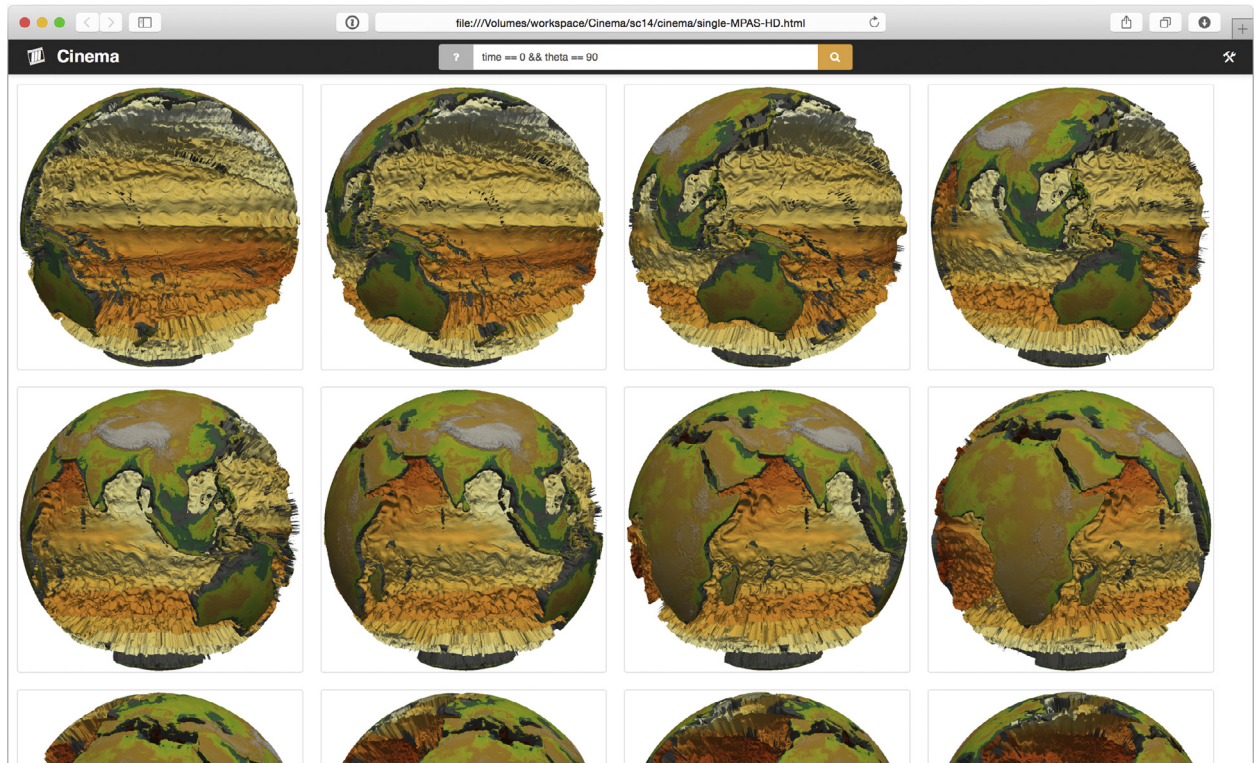


Fig. 4. The graphical user interface of the scientist-defined search pipeline for analysis and visualization object compositing.

We can automatically display multiple objects from visualization space by selecting the associated (time, objects, and camera) Cinema images and analysis products from the database and compositing them together. Cinema does not require the scientist to do this manually through a database query. Instead, the scientist uses a graphical user interface (GUI), depicted in Fig. 1(b), that emulates an application like ParaView to create an experience similar to exploring raw simulation data. In Fig. 1(b): the *Eyes* indicate analysis and visualization objects that can be interactively turned on and off; the *Plus* and *Pencil* allow for selection and coloring of specific analysis and visualization objects; and the *Time*, *Phi* and *Theta* controls allow for either selection or animation of these elements.

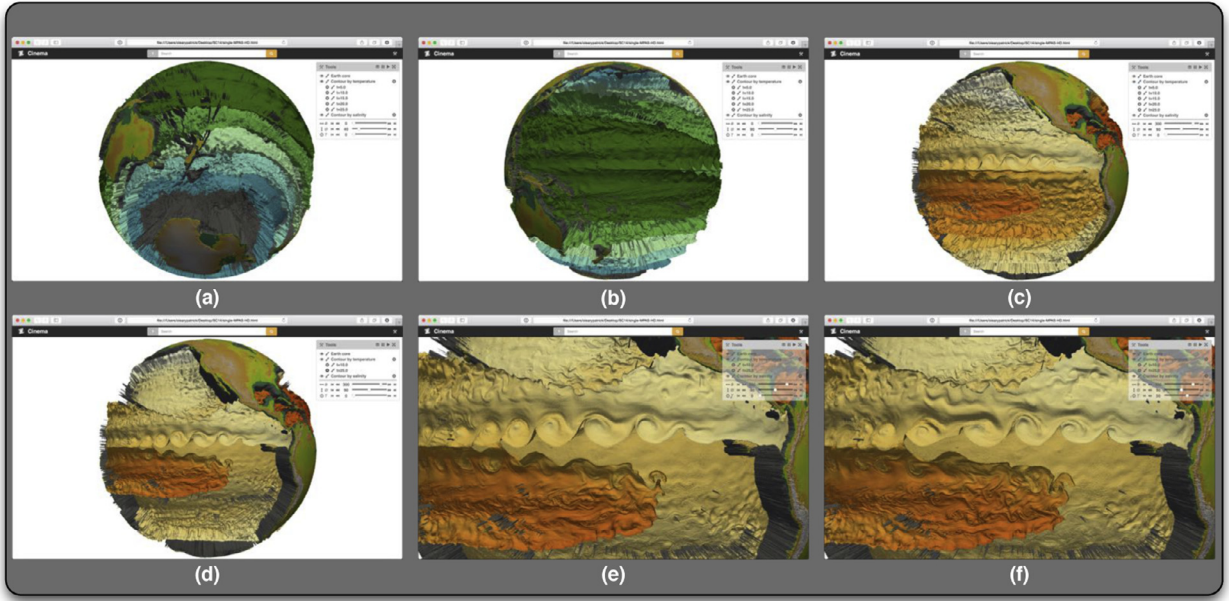


Fig. 5. (a–c) The Cinema MPAS-Ocean results using rotation about the x-, y-axis and changing to colored by salinity temperature layers. (d) Editing the composite pipeline by removing three temperature contours. (e and f) The pan and zoom interaction and the selection of different time steps.

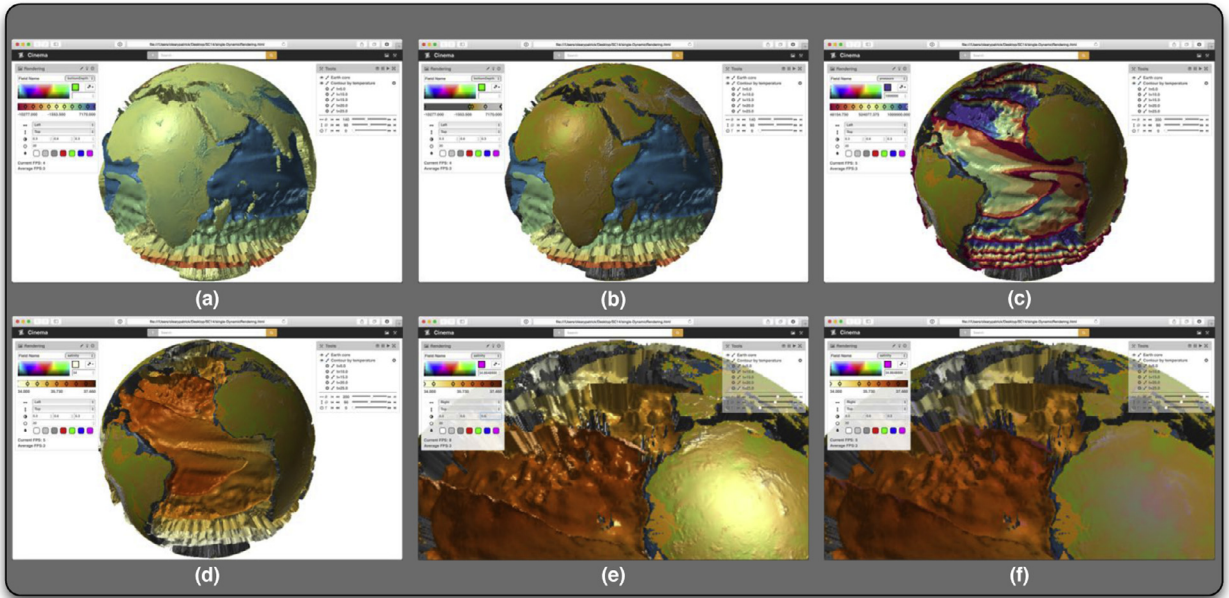


Fig. 6. The capabilities added by dynamic deferred rendering. (a and b) Temperature isosurfaces with an Earth core colored with a spectral or an Earth-like colormap. (c and d) Dynamically coloring by different parameters (pressure and salinity) and colormaps. (e and f) Increasing the spectral contribution and changing color (blue) of the light. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In addition, the dynamic rendering GUI, shown in Fig. 3(a), allows the scientist to further investigate the simulation field data associated with the analysis and visualization objects. It's these types of capabilities that pushes the Cinema experience closer to interactive exploration of raw simulation data.

3.3. Metadata and content querying

By leveraging the metadata associated with the Cinema images and analysis products in the interactive exploration database, Cinema allows the scientist to execute metadata queries (or browse analysis and visualization objects) to produce a prioritized

sequence of matching results, as demonstrated in Fig. 4. The metadata, produced by the *in situ* analysis and visualization process, includes data properties of the simulation data, such as histograms, as well as image properties.

The compositing infrastructure also makes it possible to perform queries that search on the content of the image in the database. A query could be formulated that matches on the quality of the view of a particular analysis and visualization object value. However, Fig. 4 shows a far simpler set of search results associated with a particular compositing pipeline and a query of `time == 0 && theta == 90`.

4. Results

Figs. 5 and 6 demonstrate the interactive capabilities of Cinema for visually exploring the results of the MPAS-Ocean simulation. While data reduction is highly dependent on sampling and the specified visualization pipelines, this simulation claimed an approximately 202GBs to 16GBs reduction, but an increase from 101MBs for a one image per time step traditional *in situ* workflow.

5. Conclusion

Cinema is a novel framework for image-based extreme scale *in situ* analysis and visualization. It couples *in situ* analysis and visualization outputs with an image database query method to enable interactive exploration and data browsing. The compositing pipelines provide an experience that mirrors the interactive exploration of raw simulation data, but with a constant computation time for all analysis and visualization algorithms. As implemented, Cinema meets the overarching goals for *in situ* analysis and visualization systems. It preserves important elements of simulations, while significantly reducing the data needed to preserve these elements, and offers sophisticated interactive post-processing exploration.

Acknowledgments

The authors would like to thank Patricia Fasel¹, Andrew Bauer², Benjamin Boeckel², Dave DeMarle² and Zach Mullen² for their development work involving either *in situ* MPAS-Ocean or Cinema. We would also like Francesca Samsel³ for the ocean color lookup table.

This work was funded by Dr. Lucy Nowell, [ASCR Program](#), Office of Science and ASC, Department of Energy (DOE). Patrick O'Leary, Sébastien Jourdain and Scott Wittenburg were also funded by a [DOE Office of Nuclear Energy Fast Track SBIR](#) award [DE-SC0010119](#).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.parco.2015.10.005](http://dx.doi.org/10.1016/j.parco.2015.10.005).

References

- [1] J. Ahrens, S. Jourdain, P. O'Leary, J. Patchett, D.H. Rogers, M. Petersen, An image-based approach to extreme scale *in situ* visualization and analysis, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC '14, IEEE Press, Piscataway, NJ, USA, 2014, pp. 424–434, doi:[10.1109/SC.2014.40](https://doi.org/10.1109/SC.2014.40).
- [2] J. Ahrens, S. Jourdain, P. O'Leary, J. Patchett, D.H. Rogers, P. Fasel, A. Bauer, M. Petersen, F. Samsel, B. Boeckel, In Situ MPAS-ocean image-based visualization, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, Visualization & Data Analytics Showcase, 2014 URL http://sc14.supercomputing.org/sites/all/themes/sc14/files/archive/sci_vis/sci_vis_pages/svs105.html.
- [3] T. Ringler, M. Petersen, R.L. Higdon, D. Jacobsen, P.W. Jones, M. Maltrud, A multi-resolution approach to global ocean modeling, *Ocean Model.* 69 (2013) 211–232.
- [4] J. Chen, I. Yoon, W. Bethel, Interactive, internet delivery of visualization via structured prerendered multiresolution imagery, *IEEE Trans. Vis. Comput. Graph.* 14 (2) (2008) 302–312.
- [5] A. Kageyama, T. Yamada, An approach to exascale visualization: interactive viewing of in-situ visualization, *Comput. Phys. Commun.* 185 (1) (2014) 79–85.
- [6] A. Tikhonova, C.D. Correa, K.-L. Ma, Visualization by proxy: A novel framework for deferred interaction with volume data, *IEEE Trans. Vis. Comput. Graph.* 16 (6) (2010) 1551–1559.
- [7] A. Tikhonova, H. Yu, C.D. Correa, J.H. Chen, K.-L. Ma, A preview and exploratory technique for large-scale scientific simulations, in: Proceedings of the 11th Eurographics Conference on Parallel Graphics and Visualization, EG PGV'11, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 2011, pp. 111–120.
- [8] A. Tikhonova, C.D. Correa, K.-L. Ma, An exploratory technique for coherent visualization of time-varying volume data, *Comput. Graph. Forum* 29 (3) (2010) 783–792.
- [9] K. Moreland, W. Kendall, T. Peterka, J. Huang, An image compositing solution at scale, in: Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, SC '11, ACM, New York, NY, USA, 2011, pp. 25:1–25:10.
- [10] V.S. Subrahmanian, S. Jajodia, *Multimedia Database Systems: Issues and Research Directions*, 1st ed., Springer Publishing Company, Incorporated, 2012.
- [11] J. Freire, D. Koop, F. Chirigati, C. Silva, Reproducibility using vistrails, *Implementing Reproducible Computational Research*.
- [12] N. Elmqvist, P. Tsigas, A taxonomy of 3d occlusion management for visualization, *Vis. Comput. Graph.*, *IEEE Trans.* 14 (5) (2008) 1095–1109.
- [13] J. Tao, J. Ma, C. Wang, C.-K. Shene, A unified approach to streamline selection and viewpoint selection for 3d flow visualization, *IEEE Trans. Vis. Comput. Graph.* 19 (3) (2013) 393–406.
- [14] A.H. Squillacote, D.E. DeMarle, J. Ahrens, C. Law, B. Geveci, K. Moreland, B. King, *ParaView Guide*, 1st ed., Kitware, Incorporated, 2007.

¹ Los Alamos National Laboratory

² Kitware

³ University of Texas at Austin