Computational Oceanography is Coming of Age

Thomas W. N. Haine*, Renske Gelderloos, Miguel A. Jimenez-Urias, Ali H. Siddiqui,

Earth & Planetary Sciences, Johns Hopkins University, Baltimore, MD

Gerard Lemson, Dimitri Medvedev, Alex Szalay,

Physics & Astronomy, Johns Hopkins University, Baltimore, MD

Ryan P. Abernathey,

Earth & Environmental Sciences, Columbia University, New York, NY

Mattia Almansi,

National Oceanography Centre, Southampton, UK

Christopher N. Hill

Earth, Atmospheric, & Planetary Sciences, MIT, Cambridge, MA

**Corresponding author*: Thomas Haine, Thomas.Haine@jhu.edu

ABSTRACT

Computational Oceanography is the study of ocean phenomena by numerical simulation, especially dynamical and physical phenomena. Progress in information technology has driven exponential growth in the number of global ocean observations and the fidelity of numerical simulations of the ocean in the past few decades. The growth has been exponentially faster for ocean simulations, however. We argue that this faster growth is shifting the importance of field measurements and numerical simulations for oceanographic research. It is leading to the emergence of Computational Oceanography as a branch of marine science on par with observational oceanography. Although some specific limits and challenges exist, many opportunities are identified for the future of Computational Oceanography. Most important is the prospect of hybrid computational and observational approaches to advance understanding of the ocean.

Capsule summary. Computational oceanography is an emerging discipline that asserts that high-resolution numerical ocean circulation models have the potential to be equally valuable as data from the natural ocean.

1. Introduction

The number of observations of the global ocean has grown tremendously over the last several decades. Oceanography has been transformed by this growth, particularly, knowledge of ocean circulation and dynamics, and the ocean's role in Earth's climate. Meanwhile, the fidelity of ocean general circulation models (OGCMs) has also grown tremendously. We are reaching the point that some OGCM solutions are essentially indistinguishable from observations (see Fig. 1). In the words of Ed Lorenz, we should anticipate that numerical "experiments will...duplicate the circulation to any desired degree of accuracy" (Lorenz 1967).[1] This essay explores the history of this growth and its prospects. We show that ocean observations and OGCMs have grown at different rates. We argue therefore that OGCMs are becoming equally important as ocean observations in advancing oceanography. We call this the coming of age of Computational Oceanography.

2. Unequal Exponential Growth

Two examples illustrate the growth of ocean observations. First, consider temperature observations in the global deep ocean over the last half century. Fig. 2a shows the cumulative number of temperature observations deeper than 1000m. They have grown exponentially (notice the y axis is logarithmic). Averaged over the last century, the exponential growth has a doubling period of 10.5 years, giving an approximately 50-fold expansion in the deep temperature database since 1960. Technology transitions have maintained this exponential growth, specifically, advances

[1]The prescient Lorenz was writing about atmospheric models in the late 1960s, but the message applies to OGCMs today.

in micro-electronics and information technology. In the 1990s conductivity-temperature-depth (CTD) sensors on autonomous profiling floats took over from ship CTD sampling, for example, leading in the 2000s to the transformative Argo global float network (Argo 2020).

Second, consider the history of sea level observations from satellite altimeters. Sea level data have revolutionized physical oceanography by providing information on the surface circulation, mesoscale eddies, tides, and sea level change. Fig. 2b shows the sequence of altimeter missions (coloured bars) and the cumulative number of observing days (black line). The number of observing days reveals the growth in sea level observations (although there is great variety between missions). The number of sea level observations has grown nearly exponentially since the mid 1980s with a doubling time of about 8.2 years and a $\approx$ 20-fold expansion in the sea level database since 1985. Again, micro-electronic and information technology advances have maintained this growth.

Technology advances have also fueled growth in the fidelity of OGCMs. For example, Fig. 2c shows the history of global OGCM resolution. We use the global ocean models from the Intergovernmental Panel on Climate Change (IPCC) reports and measure the model resolution with the total number of grid points. The peak resolution of the ocean OGCMs in the first IPCC report was 2.7^{o} with 9 vertical levels. The peak resolution in the latest (sixth) IPCC report is 0.067^{o} with 75 vertical levels. The growth in OGCMs is exponential with a doubling time of 2.8 years and a 1700-fold increase since 1990. For the most highly-resolved models in each assessment, the doubling time is even faster at 2.1 years.

Now compare the horizontal resolution of ocean measurements with OGCM resolution. The Argo profiling float network operates about 4000 floats at any one time. Each float makes a vertical profile from 2000m depth to the surface every ten days. The global average spacing of profiles is therefore 300km.[2] The spacing between altimeter tracks for the TOPEX/Poseidon and Jason satellite altimeters is also about 315km (at the equator), with a repeat period of ten days. The present day peak OGCM resolution of $0.067^{o} \approx 7$km is therefore 40 times higher.

3. Prospects for Future Growth

Looking ahead, the future is bright for the Argo network. The reason is that Argo is part of the Global Ocean and Global Climate Observing System, which implements the Paris Agreement on climate change and United Nations sustainable development goals. New capacities, like deep profiling floats, and new technologies, like biogeochemical sensors, are planned over the next few years (GCOS 2016). It is unclear how the network can double in size in the next decade and maintain long term exponential growth, but it is plausible.

The future is also bright for sea level measurements. The Surface Water and Ocean Topography (SWOT) mission, scheduled for launch in 2022, will start a new era of sea level observation. SWOT will observe sea level over a swath, rather than over a single patch. It will have 15km resolution, or better, covering most of the global ocean every 21 days (Morrow et al. 2019). It will improve the spatial resolution of sea level data by a factor of about ten. Therefore, the prospects for the altimetry record to continue growing exponentially in the 2020s are good.

For OGCMs, resolution improves as supercomputer technology advances. Historically, that follows Moore's "law," which says that transistor density in microprocessors doubles every two years (Moore 1975). For instance, machines first achieved petaflop speeds (10^{15} floating point operations per second) in 2008 and exaflop speeds (10^{18}) in March 2020, a doubling every 1.1 years (see Fig. 2d). On this basis, the OGCM resolution will probably continue to double every 2.8

[2] The vertical resolution of Argo profile data is about 5m, which is about 7 times higher than the best AR6 OGCMs and about 3 times higher than the Poseidon Project run mentioned below.

years, at least for another few years. It is reasonable to expect exascale IPCC OGCMs by the mid 2020s with horizontal resolutions of a few kilometers. After that, with widespread anticipation that Moore's law will end (Waldrop 2016), future growth is uncertain. Exploiting new application-specific hardware accelerators and new OGCM software architectures, like lower precision (Palem 2014), will be crucial.

4. Regime Shift

This evidence shows that information technology advances are driving exponential growth in ocean observations and exponential growth in OGCM resolution. But the OGCM growth rate is faster. Therefore, OGCM resolution is also growing exponentially faster than the growth in ocean field data. This faster growth points to a regime shift in the scientific importance of OGCMs. In 1990, OGCMs were obviously biased compared to measurements, for example, of deep temperatures or sea level. In 2020, OGCMs are achieving resolutions that are substantially greater than the gaps between measurements, at least for some regimes, like deep and abyssal ocean currents.

We should expect this trend to continue for the foreseeable future. Therefore, physical oceanography is leaving the era in which most knowledge came from observations of the real ocean. It is entering an era in which numerical circulation models are as important as observations for advancing knowledge. The ascending methodology in this new era we call **Computational Oceanography**. We argue that it is emerging as a new branch of marine science.

5. Emergence of Computational Oceanography

What are the criteria to claim that OGCM solutions should be treated, in some cases, as seriously as real measurements? Realizing them would mark the emergence of Computational Oceanography. These criteria are on our checklist:

1. **Confidence in the fidelity of the basic tools and methods.** Consider two types of tool: First, consider the theoretical definition of the ocean circulation problem. Computational Oceanography relies on software to compute approximations to the incompressible rotating-stratified Navier Stokes equations, with equations for the conservation of dissolved salts and heat (McWilliams 1996; Fox-Kemper et al. 2019). There is little doubt that these are the right equations for ocean circulation. The software is mature, stable, and diverse. The issue of unresolved processes, and parametrizing their effects remains an important area of research. For example, it is still unclear how to represent unresolved submesoscale processes on the larger scale flow. Although much progress has been made on this problem in the last 30 years (Gent 2011; Fox-Kemper et al. 2019), we believe that resolution improvements have been more important (Fig. 2d). In other words, we believe that the problem of parametrizing unresolved scales is not so pathological that it contaminates all of the resolved scales. A corollary is that OGCMs are less complicated than the real ocean, meaning that OGCM variability is a lower bound on the variability in the real system. These are *de facto* working hypotheses of all theoretical and numerical approaches to understanding the ocean circulation.

   Second, we need tools to adjust OGCM solutions to agree with observations; that is, to solve the data assimilation and state estimation problem (Bennett 1992; Wunsch 1996, 2006; Kalnay 2002). For example, state estimation is used to produce retrospective reanalyses of the time-evolving ocean state and data assimilation is used to initialize prospective forecasts of the future. Although many questions remain open, these methods are also now mature, stable, and diverse.

2. **The number of OGCM degrees of freedom exceeds the number of observational constraints.** This criterion concerns the state estimation and data assimilation problems. In essence, it is about whether it is possible (in principle) to adjust an OGCM solution to fit the observations exactly or not. If the OGCM can be adjusted to fit the data exactly, the state estimation problem is under-determined. Otherwise, it is over-determined.[3] For example, fitting a straight line through three distinct data points is over-determined because no line exists that passes through all three points (in general). But fitting a quadratic curve through two data points is under-determined because an infinite number of quadratic curves will pass through the points. The number of OGCM degrees of freedom scales as the number of grid points. The number of observational constraints scales as the number of distinct measurements. Fig. 2 shows evidence that the number of OGCM grid points per observational constraint exceeds one because, loosely, the peak OGCM resolution is now 40 times higher than the Argo and Jason data spacing. This gap is growing exponentially because OGCM resolution is growing exponentially faster than data density. Therefore, the state estimation problem is moving from (in principle) being over-determined to under-determined.

Crossing this threshold has interesting implications: First, OGCMs pass Turing or Feigenbaum tests (Turing 1950; Feigenbaum 2003; Harel 2005). That is, OGCM solutions become indistinguishable from observations of the real ocean and a subject-matter expert cannot tell them apart. In our regional OGCM simulations of the Denmark Strait Overflow (DSO) at resolutions of 0.5–2km, we are approaching this point (Fig. 1, Magaldi and Haine 2015;

[3]Ignoring the atypical case of the problem being exactly determined.

Almansi et al. 2020). Similarly, our high-resolution, regional state estimates are nearly under-determined (Lea et al. 2006; Dwivedi et al. 2011).

Second, the OGCM solutions make accurate, testable predictions about the real ocean. Historically, advances from theoretical and numerical research in dynamical oceanography have lagged advances from observational research. Once OGCMs become under-determined by data, it will be common for them to make predictions that can be tested by field programs. For example, our DSO simulations show exchange of dense water out of the overflow onto the east Greenland continental shelf, and vice versa (Magaldi et al. 2011). They also show entrainment of near-surface waters south of Iceland into the DSO within a few months, at least during hard winters (Saberi et al. 2020). It remains to be seen if these predictions occur in the real ocean.

6. Limits to Computational Oceanography

Although these opportunities are exciting, there are clear limits to Computational Oceanography. Specifically, direct numerical simulation (DNS) of the global ocean circulation is inconceivable today. DNS in this context means running OGCMs that resolve all scales of motion; from the planetary scale to the dissipation scale (around 1mm), and from centuries to seconds. DNS would avoid the challenge of parametrizing the effects of the unresolved scales, but at vast computational cost. Fig. 3 shows why. It shows the full range of space and time scales relevant to the ocean general circulation, about ten orders of magnitude in both. It also shows the space time volumes accessible to present-day supercomputers, including the best AR6 OGCMs shown in Fig. 2, the Poseidon Project run, and turbulence simulations (DNS and large eddy simulations, LES). To span the entire space time plane, super computers would need to resolve about 10^{25} grid points and 10^{10} time steps. That is about 16 orders of magnitude more grid points than is possible today.

Extrapolating the doubling time of 2.8 years in Fig. 2c, it would take 150 years to achieve this increase, which is impossible to envision. Clearly, OGCM simulations cannot replace observations of the natural ocean.

7. Opportunities for Computational Oceanography

The opportunities for Computational Oceanography to advance marine science include:

- Migration from the study of specific instances of phenomena to the study of statistics of these phenomena. The DSO is one of many currents that is affected by rotation, stratification, and bathymetry. It is inconceivable to observe all of them, but they can all be simulated in an exascale OGCM. Empirical characterization of these numerical overflows would be an important step forward.

- Discovery and characterization of intermittent, time-dependent, three-dimensional phenomena, which are hard to observe. Submesoscale currents are in this class, which occur at horizontal scales shorter than several kilometers (Thomas et al. 2008). Diapycnal mixing is another example, which occurs at scales shorter than meters (MacKinnon et al. 2017).

- Comprehensive and illuminating analyses of ocean mass, heat, salt, momentum, energy, and vorticity budgets, in a way that is nearly impossible with direct observations.

- Discovery and characterization of ocean circulation regimes that cannot be observed. Examples include the circulation during the last glacial maximum (paleo-oceanography) or in extra-terrestrial oceans (exo-oceanography). For these ocean circulation problems, the data-sparseness challenge is much worse than for the modern ocean (LeGrand and Wunsch 1995; Amrhein et al. 2018; Way et al. 2017). Criterion 2 was achieved with smaller computational resources for these fields, and therefore they have already entered the era of Computational Oceanography by the rationale in section 5.

- Robust observing system design using OGCM solutions as synthetic data. These Observing System Simulation Experiments (Errico et al. 2012) should become the best-practice standard for fieldwork design. There are implications for making the OGCM output accessible and easy to work with (see below), but the payoff from engaging observational oceanographers is great.

- Insight from OGCM state estimation to support fieldwork, ideally in real time. The community should embrace the fact that the under-determined state estimates imply an infinite number of OGCM solutions that match the data exactly. This means that techniques are needed to characterize and handle the OGCM null space (indeterminacy). For example, observational oceanographers at sea could make decisions about where, when, and how to observe using OGCM information that captures the range of possible circulation states consistent with data. This practice is common in atmospheric science already.

- More efficient identification of interesting phenomena using automatic methods, like artificial intelligence and data mining (Kutz 2017; Lguensat et al. 2019). In fact, such automatic methods will become essential as the size of OGCM output grows exponentially and overwhelms manual feature identification (see below).

- Increasing transition of dynamical oceanography to an experimental (computational) science. It has long been recognized that idealized models isolate physical mechanisms relevant to the general circulation and thereby build dynamical understanding. We still require idealized models; in particular, we need a hierarchy of models that span the gap between geophysical fluid dynamics problems and realistic simulations of the circulation. This hierarchy will ensure that the increasing OGCM realism does not outpace understanding of the basic physics (Held 2005; Vallis 2016; Emanuel 2020).

8. Prospects for Computational Oceanography

How can these priorities be achieved and what are the prospects for Computational Oceanography? We should focus on these issues in the next several years:

- OGCM simulation output must be "democratized" to lower barriers to dissemination. The output should be freely available, including to non-professional users. Traditionally, effort has focused on the challenges of calculating OGCM solutions with supercomputers. The OGCM output has become increasingly harder to use, because of the massive data volume, and the technical complexities that attend the high-performance computation. Access to high-resolution OGCM output is restricted to a few experts in practice.

  The remedy is to build high-performance data science infrastructure to match the high-performance compute infrastructure. These data portals must be open and have low thresholds to getting started. We must be able to sample the simulations the way that we sample the real ocean. For example, it should be easy for an observational oceanographer to plot a synthetic hydrographic section or mooring timeseries. The data portals must include open software and significant compute resources to process and analyze the simulation data. We must avoid the inefficient practice in which users are forced to download voluminous data to their local machines and then write their own code to analyze them. Technologies and infrastructure to achieve these goals are under development, such as the OceanSpy OGCM data analysis package (Almansi et al. 2019), the Pangeo community in geoscience big data (`pangeo.io`), the SciServer big data science platform (Medvedev et al. 2016), and other cloud analysis clusters.

- "Benchmark" OGCM reference solutions should be computed using the best available compute resources and served to the public. For example, we are computing global OGCM solutions at (nominally) 1km horizontal resolution as part of the Poseidon Project (`poseidon.idies.jhu.edu`).[4] Such benchmark solutions are of intrinsic value to all oceanographers, not just ocean modellers, for the reasons stated above. Benchmark solutions for regional ocean circulation problems are valuable for the same reasons, as are idealized simulations of specific ocean dynamical processes. The track record of other fields using this approach is impressive. For instance, the Johns Hopkins Turbulence Database exposes large-scale turbulent data to researchers and provides easy-to-use interfaces to retrieve and interact with the data using novel metaphors like immersing virtual sensors into the 4-D data (`turbulence.pha.jhu.edu`; Perlman et al. 2007; Li et al. 2008).

- OGCMs will migrate to exascale compute resources in the next few years. This migration will involve new paradigms to access the data. For example, with today's petaflop supercomputers only about 0.1% of the OGCM solution can be permanently stored for analysis. The problem arises because of the prohibitive time needed to transfer the massive output volume to long-term storage media, and the prohibitive expense of the media. This loss of OGCM data will be much worse on exaflop machines.

  To mitigate this problem consider the strategy adopted by the Large Hadron Collider (LHC), the world's most sophisticated experimental facility. The LHC provides a single source of data on subatomic particle collisions. Several experiments tap the data stream in so-called "beam-lines." Within each experiment, customized hardware monitors the stream. Only about one event in ten million is retained for storage and detailed analysis. In exascale oceanography the analogous idea (see section 3.3.5 in Asch et al. 2018) is to enable automatic identification of

[4] The Poseidon Project is unrelated to the TOPEX/Poseidon altimeter.

selected circulation events and trigger storage while the OGCM runs. For example, we could target intermittent intense mixing events, plus their antecedents and fates. An implication is that we must build a software interface for community-supplied software plugins to implement the custom triggers. Also, we need to enable posterior re-computation of small space-time chunks of the full solution, with customized diagnostics, and possibly at higher resolutions.

Computational Oceanography promises powerful new tools to address previously intractable problems. It does not aim to supplant observational oceanography. Indeed, observing the natural ocean must never cease. Instead, the greatest opportunity lies in merging these hitherto disparate branches of our field. Lasting progress will require that we trust computational insights, verify them with real world observations, and understand them with fundamental theory.

Acknowledgments. This material is based upon work supported by the National Science Foundation under Grant Number OAC-1835640 and by the Institute for Data Intensive Engineering and Science at Johns Hopkins University.

Data availability statement. Codes to make the figures are available at `github.com/hainegroup/Computational-Oceanography-Commentary`. Fig. 1 is based on model and field data in Haine (2010). For Fig. 2, the temperature data are from the National Centers for Environmental Information World Ocean Database, the altimeter mission data are from `www.altimetry.info`, the IPCC data are from the IPCC reports and `pcmdi.llnl.gov/CMIP6`, and the supercomputer data are from `en.wikipedia.org/wiki/List_of_fastest_computers`. The data for AR6 are from the HighResMIP project in July 2020, which was incomplete then. Fig. 3 is based on Klinger and Haine (2019).

References

Almansi, M., R. Gelderloos, T. W. N. Haine, A. Saberi, and A. H. Siddiqui, 2019: OceanSpy: A Python package to facilitate ocean model data analysis and visualization. *Journal of Open Source Software*, **4 (39)**, 1506, doi:10.21105/joss.01506.

Almansi, M., T. W. N. Haine, R. Gelderloos, and R. S. Pickart, 2020: Evolution of Denmark Strait Overflow cyclones and their relationship to overflow surges. *Geophys. Res. Lett.*, doi: 10.1029/2019GL086759.

Amrhein, D. E., C. Wunsch, O. Marchal, and G. Forget, 2018: A global glacial ocean state estimate constrained by upper-ocean temperature proxies. *J. Climate*, **31 (19)**, 8059–8079, doi: 10.1175/jcli-d-17-0769.1.

Argo, 2020: Argo float data and metadata from global data assembly centre (Argo GDAC). SEANOE, doi:10.17882/42182.

Asch, M., and Coauthors, 2018: Big data and extreme-scale computing. *The International Journal of High Performance Computing Applications*, **32 (4)**, 435–479, doi:10.1177/1094342018778123.

Bennett, A. F., 1992: *Inverse methods in physical oceanography*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 346 pp., doi:10.1017/cbo9780511600807.

Dwivedi, S., T. W. N. Haine, and C. E. Del Castillo, 2011: Upper ocean state estimation in the Southern Ocean Gas Exchange Experiment region using the four–dimensional variational technique. *J. Geophys. Res.*, **116, C00F02**, doi:10.1029/2009JC005615.

Emanuel, K., 2020: The relevance of theory for contemporary research in atmospheres, oceans, and climate. *AGU Advances*, **1 (2)**, doi:10.1029/2019av000129.

Errico, R. M., R. Yang, N. C. Privé, K.-S.Tai, R. Todling, M. E. Sienkiewicz, and J. Guo, 2012: Development and validation of observing-system simulation experiments at NASA's global modeling and assimilation office. *Q. J. R. Meteorol. Soc.*, **139 (674)**, 1162–1178, doi: 10.1002/qj.2027.

Feigenbaum, E. A., 2003: Some challenges and grand challenges for computational intelligence. *Journal of the ACM*, **50 (1)**, 32–40, doi:10.1145/602382.602400.

Fox-Kemper, B., and Coauthors, 2019: Challenges and prospects in ocean circulation models. *Frontiers in Marine Science*, **6**, doi:10.3389/fmars.2019.00065.

GCOS, 2016: The global observing system for climate. Tech. Rep. GCOS No. 200, World Meteorological Organization. https://library.wmo.int/index.php?lvl=notice_display&id=19838#.XxWxPJNKhBw.

Gent, P. R., 2011: The Gent-McWilliams parameterization: 20/20 hindsight. *Ocean Modelling*, **39**, 2–9, doi:10.1016/j.ocemod.2010.08.002.

Haine, T. W. N., 2010: High-frequency fluctuations in Denmark Strait Overflow transport. *Geophys. Res. Lett.*, **37**, L14 601, doi:10.1029/2010GL043272.

Harel, D., 2005: A Turing-like test for biological modeling. *Nature Biotechnology*, **23 (4)**, 495–496, doi:10.1038/nbt0405-495.

Held, I. M., 2005: The gap between simulation and understanding in climate modeling. *Bull. Amer. Meteor. Soc.*, **86 (11)**, 1609–1614, doi:10.1175/bams-86-11-1609.

Kalnay, E., 2002: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/cbo9780511802270.

Klinger, B. A., and T. W. N. Haine, 2019: *Ocean Circulation in Three Dimensions*. 1st ed., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, URL http://www.cambridge.org/9780521768436.

Kutz, J. N., 2017: Deep learning in fluid dynamics. *J. Fluid Mech.*, **814**, 1–4, doi:10.1017/jfm.2016.803.

Lea, D. J., T. W. N. Haine, and R. F. Gasparovic, 2006: Observability of the Irminger Sea circulation using variational data assimilation. *Q. J. R. Meteorol. Soc.*, **132 (618)**, 1545–1576, doi:10.1256/qj.05.77.

LeGrand, P., and C. Wunsch, 1995: Constraints from paleotracer data on the North Atlantic circulation during the Last Glacial Maximum. *Paleoceanography*, **10 (6)**, 1011–1045, doi:10.1029/95pa01455.

Lguensat, R., J. Le Sommer, S. Metref, E. Cosme, and R. Fablet, 2019: Learning generalized quasi-geostrophic models using deep neural numerical models, 1911.08856v1.

Li, Y., and Coauthors, 2008: A public turbulence database cluster and applications to study lagrangian evolution of velocity increments in turbulence. *Journal of Turbulence*, **9**, N31, doi:10.1080/14685240802376389.

Lorenz, E. N., 1967: *The Nature and Theory of the General Circulation of the Atmosphere*. WMO: TP, World Meteorological Organization.

MacKinnon, J. A., and Coauthors, 2017: Climate Process Team on internal-wave driven ocean mixing. *Bull. Amer. Meteor. Soc.*, **98 (11)**, 2429–2454, doi:10.1175/bams-d-16-0030.1.

Magaldi, M. G., and T. W. N. Haine, 2015: Hydrostatic and non-hydrostatic simulations of dense waters cascading off a shelf: the East Greenland case. *Deep Sea Res., Part I*, **96**, 89–104, doi:10.1016/j.dsr.2014.10.008.

Magaldi, M. G., T. W. N. Haine, and R. S. Pickart, 2011: On the nature and variability of the East Greenland Spill Jet: A case study in summer 2003. *J. Phys. Oceanogr.*, **41**, 2307–2327, doi:10.1175/JPO-D-10-05004.1.

McWilliams, J. C., 1996: Modeling the oceanic general circulation. *Ann. Rev. Fluid Mech.*, **28**, 215–248, doi:10.1146/annurev.fl.28.010196.001243.

Medvedev, D., G. Lemson, and M. Rippin, 2016: SciServer compute. *Proceedings of the 28th International Conference on Scientific and Statistical Database Management - SSDBM '16*, ACM Press, doi:10.1145/2949689.2949700.

Moore, G. E., 1975: Progress in digital integrated electronics IEDM Tech Digest. IEEE.

Morrow, R., and Coauthors, 2019: Global observations of fine-scale ocean surface topography with the surface water and ocean topography (SWOT) mission. *Frontiers in Marine Science*, **6**, doi:10.3389/fmars.2019.00232.

Palem, K. V., 2014: Inexactness and a future of computing. *Phil. Trans. R. Soc. Lond. A*, **372 (2018)**, 20130 281, doi:10.1098/rsta.2013.0281.

Perlman, E., R. Burns, Y. Li, and C. Meneveau, 2007: Data exploration of turbulence simulations using a database cluster. *Proceedings of the 2007 ACM/IEEE conference on Supercomputing - SC '07*, ACM Press, doi:10.1145/1362622.1362654.

Saberi, A., T. W. N. Haine, R. Gelderloos, M. F. de Jong, H. Fury, and A. Bower, 2020: Lagrangian perspective on the origins of Denmark Strait Overflow. *J. Phys. Oceanogr.*, **50 (8)**, 2393–2414, doi:10.1175/JPO-D-19-0210.1.

Thomas, L. N., A. Tandon, and A. Mahadevan, 2008: Submesoscale processes and dynamics. *Ocean Modeling in an Eddying Regime*, American Geophysical Union, 17–38, doi:10.1029/177gm04.

Turing, A. M., 1950: I. Computing machinery and intelligence. *Mind*, **LIX (236)**, 433–460, doi:10.1093/mind/lix.236.433.

Vallis, G. K., 2016: Geophysical fluid dynamics: whence, whither and why? *Proc. R. Soc. Lond. A*, **472 (2192)**, 20160 140, doi:10.1098/rspa.2016.0140.

Waldrop, M. M., 2016: The chips are down for Moore's law. *Nature*, **530 (7589)**, 144–147, doi:10.1038/530144a.

Way, M. J., and Coauthors, 2017: Resolving orbital and climate keys of earth and extraterrestrial environments with dynamics (ROCKE-3d) 1.0: A general circulation model for simulating the climates of rocky planets. *The Astrophysical Journal Supplement Series*, **231 (1)**, 12, doi:10.3847/1538-4365/aa7a06.

Wunsch, C., 1996: *The ocean circulation inverse problem*. 1st ed., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/cbo9780511629570, 442 pp.

Wunsch, C., 2006: *Discrete Inverse and State Estimation Problems*. 1st ed., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/cbo9780511535949, 371 pp.

LIST OF FIGURES

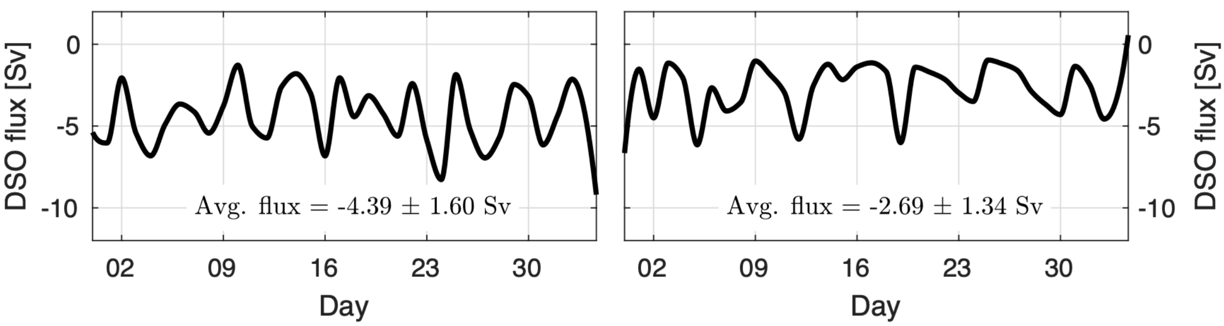

Fig. 1. OGCM Turing test. The two timeseries show Denmark Strait Overflow (DSO) volume flux (Sv, 1Sv = $10^6 \mathrm{m}^3\mathrm{s}^{-1}$, negative means equatorwards). The DSO is an important deep current between Greenland and Iceland. One timeseries is from *in situ* measurements, the other is from a high resolution OGCM. The question is: which is which? The timeseries are different, but the difficulty in answering the question reflects the small systematic error in the OGCM and therefore its realism. The two timeseries have been processed similarly. Adapted from Haine (2010).

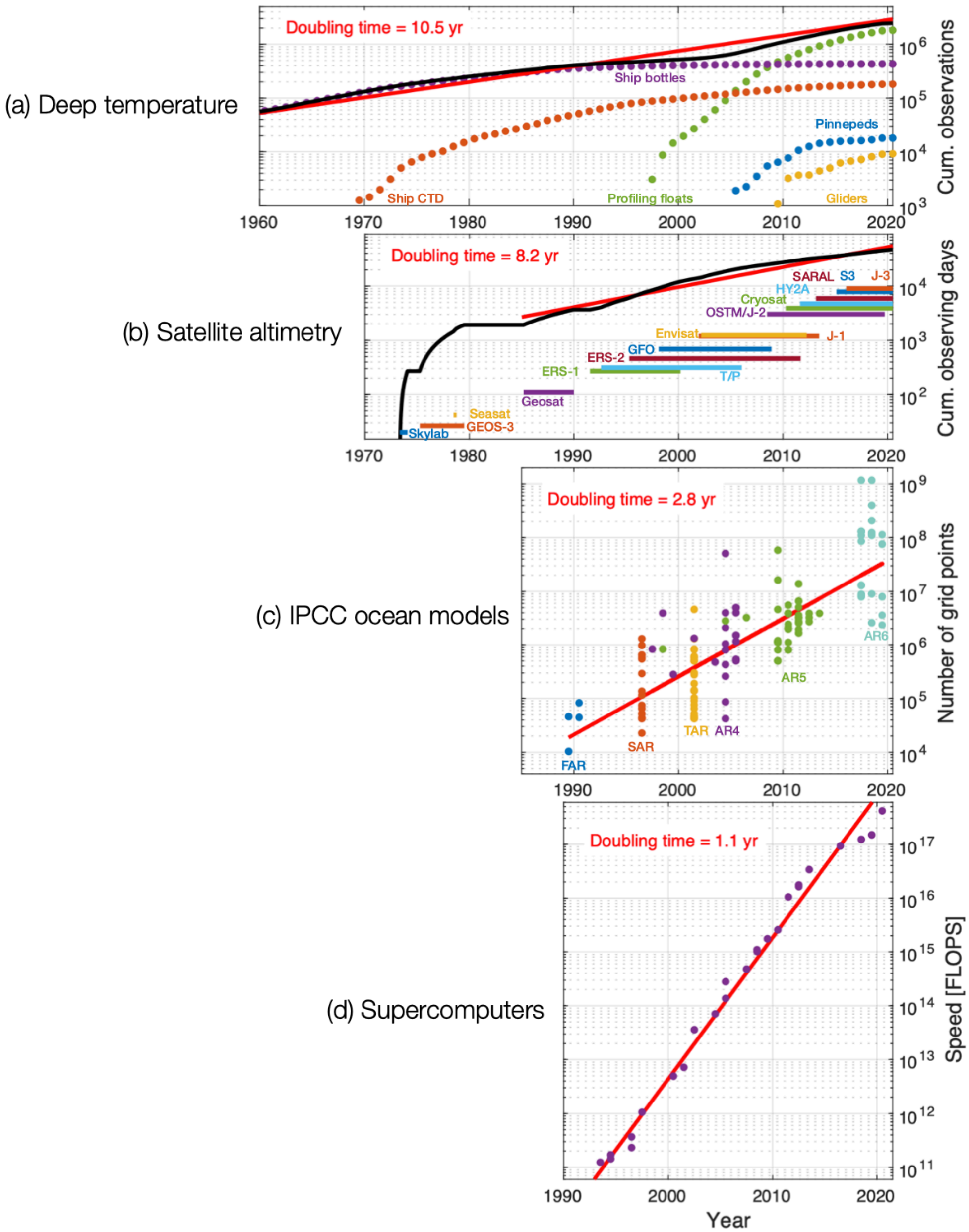

Fig. 2. Unequal exponential growth. (a) History of deep (deeper than 1000 m) ocean temperature measurements. The coloured dots show different instruments and observing platforms. (b) History of sea level measurements from satellite altimetry expressed by the cumulative number of days of measurement. The satellite missions and their durations are indicated with the coloured bars. (c) History of IPCC ocean model resolution expressed by the number of model grid points. Each coloured dot represents one ocean model: the colour indicates the IPCC assessment. (d) History of top supercomputer speeds (FLOPS = floating point operations per second). The red lines show best fit exponential growth in each panel.

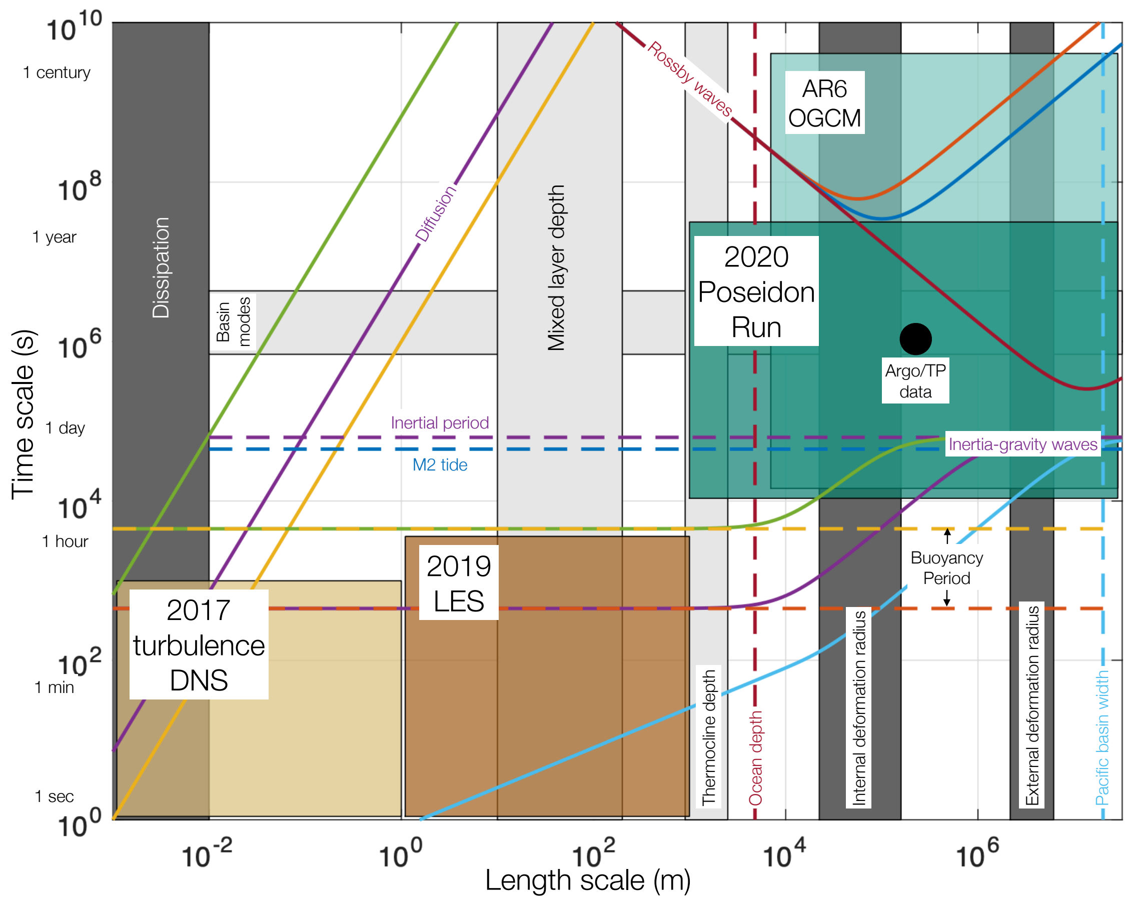

FIG. 3. Characteristic space and time scales of the ocean general circulation. Various geophysical and theoretical scales are shown with gray patches and coloured lines (for a discussion, see Klinger and Haine 2019). The four coloured rectangles show state-of-the-art circulation models (direct numerical simulation of turbulence, large eddy simulation, the Poseidon Project run and AR6 HighResMIP OGCMs). The black dot shows the sampling characteristics of the Argo profiling floats, and the TOPEX/Poseidon-Jason altimeters. The diagram is indicative, not definitive, because it suppresses the anisotropies and inhomogeneities present in the general circulation.