

Property Testing for Ocean Models. Can We Specify It?

Deepak A. Cherian

Golden, Colorado

deepak@cherian.net

I take inspiration from the property-testing literature, particularly the work of Prof. John Hughes [17], and explore how such ideas might be applied to numerical models of the ocean. Specifically, I ask whether geophysical fluid dynamics (GFD) theory, expressed as property tests, might be used to address the oracle problem of testing the correctness of ocean models. I propose that a number of simple idealized GFD problems can be framed as property tests. These examples clearly illustrate how physics naturally lends itself to specifying property tests. Which of these proposed tests might be most feasible and useful, remains to be seen.

1 Introduction

1.1 Background

Simulated possible projections of the Earth’s climate by Earth System Models are an important tool in understanding, adapting to, and mitigating the effects of climate change. Such models bring together a number of individual component models (e.g. the atmosphere, ocean, land, rivers, and more) coupled to each other through modeled interactions at the interfaces between these components (for example, the sea surface). The complexity of each individual component is vast, the complexity of a full Earth System Model is staggering (Figure 1).

Consider the ocean. The ocean is a thin shell of fluid on the Earth¹. The ocean is “stratified”, meaning the density of seawater varies with depth so that lighter fluid overlays denser fluid. Commonly, warm and fresh (low salinity) water overlays colder, saltier water. The equations that describe the motion of the ocean are the Navier-Stokes equations specialized for rotating fluids. These equations express the

¹compare its depth of ~ 5 km to the radius of the earth (~ 6400 km)

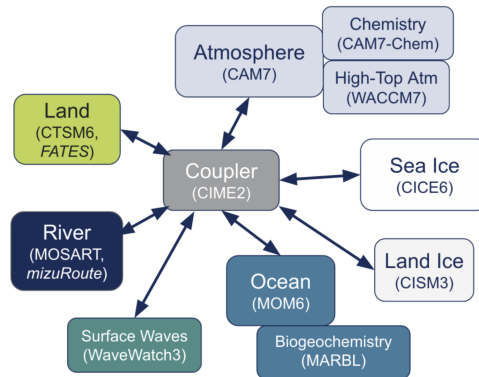


Figure 1: Schematic of the Community Earth System Model version 3 (CESM3) [20].

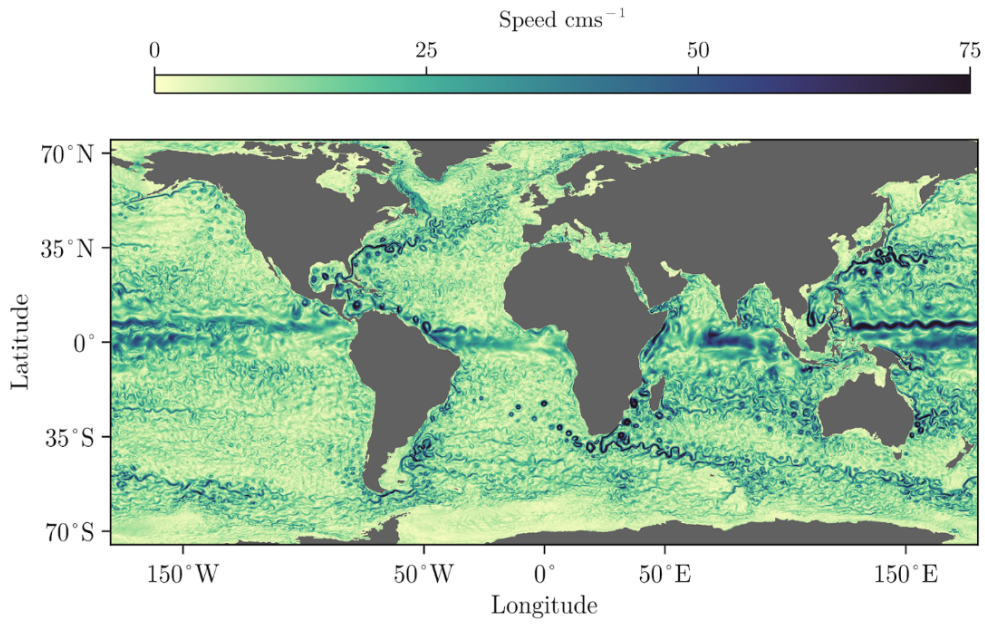


Figure 2: Sea surface speed from a global 1/12° simulation using the Oceananigans model [32]. Note several properties of the ocean circulation: (a) faster currents at western continental boundaries of ocean basins, (b) the east-west orientation of flows in the Tropics away from continental boundaries, and (c) the presence of coherent vortices (circular features).

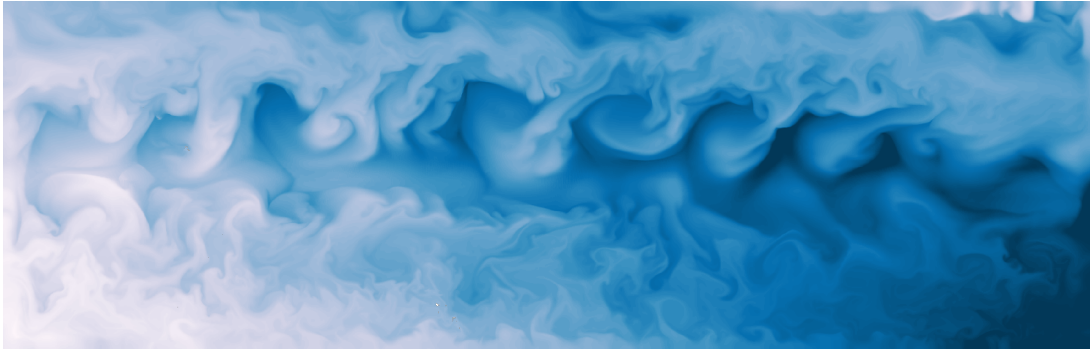


Figure 3: Sea surface temperature in the Tropical Pacific from a regional MIT General Circulation Model simulation using a $1/20^\circ \approx 5$ km spacing [6]. Darker colors indicate colder water. The full domain spans 20 degrees of latitude and 80 degrees of longitude. The complex and turbulent nature of the flow is evident.

conservation of momentum in three dimensions (x , y , and z) for three velocity components (u , v , and w); conservation of mass; conservation of heat and salt; and a thermodynamic “equation of state” that relates density ρ to temperature T and salinity S . A similar set of equations can be written down for the atmosphere. The existence of unique solutions for this set of equations is unproven [11]. Solutions for the case of geophysical fluids are further complicated by the presence of complex boundary conditions. For the ocean this includes the oceanic boundary with the landmass at its side, bottom topography, and the surface boundary with the atmosphere across which both kinetic energy and heat flow. Flows in both the ocean and atmosphere are complex and turbulent. The range of motions in the ocean range from the “basin-scale”, approximately 5000 km, to the smallest scales where molecular viscosity and diffusivity take charge, approximately a micrometer. Given the computational expense, global-scale models tend to solve the equations of motion of grids with spacing $O(10\text{ km})$ – $O(100\text{ km})$ or so. The consequences of physics at scales smaller than the grid spacing is represented using crude approximate models called “parameterizations”. Common parameterizations model the effect of turbulence, waves, clouds and other phenomena too small to be simulated directly. The complexity of the system is immense, and the computational expense of solving these equations is large. Figures 2 and 3 illustrate the complexity of modeled ocean flows using surface speed and surface temperature.

Similar observations of complexity extend to all other components of an Earth System Model (Figure 1). Projections constructed using these models form part of the scientific backing for the reports of the Intergovernmental Panel on Climate Change (IPCC). These models also enable experimentation and the creation of new knowledge by allowing scientists to tease apart relationships that are too complex to be gleaned by analytical study of the equation set. It is essential that these models be as correct as we can make them.

1.2 Testing climate models

The testing of climate model components suffers from the “oracle” problem — given a numerically constructed solution for the evolution of a system from an initial state, there is no “oracle” to tell us whether the solution is correct. When configured to match the real ocean, we might use observations to judge the quality of an integration, but this approach suffers from the problem of “compensating errors” which may hide real bugs. Unsurprisingly, testing methodologies for ocean/atmosphere models therefore rely heavily on regression tests, and comparison to reference solutions. Regression testing is treated strictly, with developers emphasizing bitwise reproducibility when comparing to reference solutions or solutions from an earlier version of the model. Reference solutions to a somewhat arbitrary selection of initial value problems for the atmosphere have been published [12, 14, 29, 38]. Interestingly [38] provide error metrics and recommend constructing specific plots for visual inspection as a testing strategy.

A quick survey of ocean modeling systems highlights the similarity in approach to testing. The MIT General Circulation Model (MITgcm) project provides reproducible test examples with reference output for comparison [25]. Similarly the Regional Ocean Modeling System (ROMS; [31]), the Coastal and Regional Ocean Community Model (CROCO; [8]), and the Modular Ocean Model (MOM6, [26]) all take a similar approach. Intercomparison with models is another approach to correctness; for example [15] report comparing the Finite-Volume Coastal Ocean Model (FVCOM) against ROMS. Other testing approaches common to many modeling systems include testing that output from a serial and parallel integration are the same, and that a simulation with no restarts is identical to one with a restart in the middle.

The developers of the MOM6 system employ a two unique testing strategies — termed “novel tests” [27]. One is to assert dimensional consistency of the code, after having hand-coded physical units for

terms in the codebase. The second is to test that simulations are invariant to a rotation of the domain by some number of 90° turns. These “novel tests” are property tests in all but name.

More recently, lightweight formal methods are beginning to be applied to sub-components of these models. For example, Altuntas and Baugh [1] use a hybrid theorem prover to test for bugs in a widely used ocean model parameterization (K-Profile Parameterization; KPP, [19]). Altuntas et al (this volume) systematically argue for greater use of formal methods across all sub-components of a climate model systems.

1.3 Generative property-based testing

One solution to the “oracle” problem is to use generative property-based testing as popularized by the Haskell library QuickCheck [7]. In this approach, the programmer specifies properties of the System Under Test, and designs a generator for valid inputs to the test. Outputs derived from a wide range of generated inputs are then tested for correctness using these properties. . The popularity of such property tests is rising particularly in the form of “metamorphic testing”, where a “metamorphic relation” describes the relationship between outputs generated from related inputs (for example, [5]). In the scientific Python ecosystem, the Hypothesis library [21] is increasingly used for testing software used for scientific analysis, including that of climate model output. For example, nascent suites of property and stateful tests are being built in the Xarray [40] and Zarr [41] scientific Python projects.

This paper is a speculative exploration of how we might apply property-based testing ideas to the testing ocean modeling systems, guided by experience building such test suites for Xarray and Zarr. We are led to ask: “what are the properties of the ocean circulation?” and are inevitably led to the field of “geophysical fluid dynamics”.

1.4 Geophysical Fluid Dynamics (GFD)

The complexity of geophysical fluid systems does not only affect numerical approaches. It took nearly 300 years of analytic effort to write down the modern formulation of the equations of motion [10]. Over that time and since then, much work has been done to understand these equations in much simpler limits that are amenable to theoretical exploration. The broad field of geophysical fluid dynamics (GFD; [35]) is the study of fluid motion governed by these equations (specialized to the Earth or other planets), with the goal of elucidating fundamental physical principles and relationships. Geophysical fluid dynamicists have studied flows in a hierarchy of complexity from the shallow water equations (constant density fluid), to quasi-geostrophic flow (slow flows in large domains with relatively large density variations in the vertical), to primitive equation models such as those used to construct the solutions in Figures 2 and 3. Solutions to equations with varying levels of approximation, and under specific cases, are available in the textbooks [13, 28, 36]. Below, I use these simplified cases and explore how they may be framed as property tests. These simplified GFD problems use simpler domains (e.g. a flat bottom or a square basin) and commonly apply approximations to the representation of the effect of the Earth’s rotation, and the complexity of vertical variations in density. The former is a recurring theme in the rest of the manuscript, so I discuss it briefly.

1.4.1 Representing a rotating planet

The influence of the Earth’s rotation in our rotating frame of reference is expressed through the Coriolis force terms $-fv$ and fu in the x and y momentum equations, where $f = 2\Omega \sin \phi$ [s⁻¹] where Ω is the

angular velocity of the Earth and ϕ is the latitude [30]. Thus the effect of Earth’s rotation on fluid flow varies with latitude, with its magnitude being negligible at the equator ($\phi = 0$), and maximum at the poles ($\phi = 90^\circ$). A truncated Taylor series expansion of $\sin \phi$ is used in simpler models that cover limited spans in latitude $f = f_0 + \beta y$. Two common limits are the f -plane where $\beta = 0$, and the β -plane where $\beta \neq 0$. On the f plane there is no difference between the latitude and longitude directions. On the β plane, oceanic motions are preferentially oriented along lines of constant latitude (“zonal”), and there is an east-west asymmetry due to the physics of a new class of wave — the Rossby wave. Both of these *properties* will be used later.

2 Dynamical property tests for ocean models

The most evident property of the ocean circulation is the east-west asymmetry: flows at western boundaries are more energetic (Figure 2). Yet testing the ability to reproduce this property simply tests that the code is able to simulate a rotating spherical planet. Surely this property is too simple! How might we do better?

Prof. John Hughes lays out a guide with 5 guiding suggestions for thinking about property tests in a paper [17] and in following talks (for example, [18])². Here I follow those suggestions and ask how might we write property tests for ocean models. As an exercise I constrain myself to writing out dynamical properties for the *whole* system, albeit in simple limits, rather than focus on individual terms or subroutines that comprise an ocean model. This is not to suggest that such tests are most likely to be useful at finding bugs. Instead this constraint simply makes the exercise more fun and interesting (to me). Though the discussion is centered on oceanography, nearly all of it can be easily translated to atmospheric and other geophysical fluid models.

In writing out these ideas I find that

1. many of these ideas are already used as either example problems or regression tests in the current generation of ocean models; and that
2. the physics of the underlying system clearly yields several dynamical properties. Yet it is not obvious which might be most useful in finding bugs.

Now I address each of Hughes’ suggestions in turn. For each idea, I describe a simple property test for the problem of sorting a list of integers to motivate the thinking behind the proposed property tests that follow.

2.1 Is there an invariant?

Example: Sorting is idempotent so sorting a sorted list yields the original sorted list.

2.1.1 Conservation of properties

Physical systems naturally have integral invariants namely the conservation of mass (or volume), energy, angular momentum, and “potential vorticity” [24]. Given that, one might configure a set of arbitrary initial conditions for the ocean state, integrate the model forward for n timesteps, with n chosen to limit computational expense. At the end of the integration, conservation principles must be satisfied to the level expected from the numerics of the model. However totally arbitrary initial conditions are not

²I acknowledge inspiration from at least two other authors: Hillel Wayne [37] and Sean Wlaschin [39].

possible, they will need to be carefully designed to prevent model blowup. It seems more useful to test such invariants in the context of other more physically motivated tests below.

2.1.2 Symmetries

The Navier-Stokes equation set has underlying symmetries. The standard example is “Galilean invariance” which means the solution is invariant to switching to a reference frame with a constant translating velocity. Other symmetry groups include rotational symmetry (currently used in the MOM6 “novel tests”), and a more general “scale invariance” [22, their equations 1.9–1.12]. These invariants are all applicable to the specific solutions used in the tests described below.

2.1.3 Balanced flows remain balanced

A more dynamically motivated invariant, in the spirit of this article, is the idea that “*balanced flows must remain balanced*”.

When a flow is *balanced* this means that the velocity and pressure fields are related functionally, and that one is in principle derivable from the other [23]. Crucially, a *lack of balance* is characterized by freely propagating waves (*transients*) that complicated the solution dearly. For ease of comparison, we will need to design tests that result in either no transients, or known transients. In this section, we initialized the model with a balanced flow. If truly balanced to the level that the numerics allow, the transients must be negligible. Further we can test our specification of balanced flows using property tests too! The classic roundtrip property test applies: by definition, transforming from velocity to pressure and back, or pressure to velocity and back, must yield the initial field.

The simplest example here is “geostrophic balance” — the balance between the pressure gradient force and the Coriolis force:

$$fu = -\frac{1}{\rho} \frac{\partial p}{\partial y}; \quad fv = \frac{1}{\rho} \frac{\partial p}{\partial x} \quad (1)$$

For arbitrary choices of f and u, v (say), there exists a pressure field p that balances that u, v . If initialized with such a balanced field, the model must maintain this balance. Classic choices for v, p are circular vortices or jets. When f is a constant, there is no dynamical difference between the x and y directions so the model must be able to maintain jets oriented at arbitrary angles. Other formulations of balanced flows add density stratification — “thermal wind balance”; and nonlinear advection — “cyclotrophic balance”.

For more complexity, we turn to the idea of balanced waves to bring in the time derivative term. Many types of balanced waves exist in all approximations to the governing equations, all of these would make good tests. Examples include surface gravity waves at the air-sea interface familiar from visits to the beach, internal gravity waves (similar to surface waves but propagate at internal density interfaces within the fluid), and much larger and slower Rossby waves that can exist in rotating fluid systems. The waves that exist for a system are described by the *dispersion relation* which relates the wavenumbers and frequencies of waves that can exist given a chosen domain. For example, internal waves in a rotating stratified fluid satisfy the dispersion relation

$$\omega^2 = \frac{f^2 m^2 + N^2 (k^2 + l^2)}{k^2 + l^2 + m^2} \quad (2)$$

where ω is the frequency and (k, l, m) are the three components of the wavenumber vector. Exact solutions can be written depending on the chosen analytic form of N and the boundary conditions. If

initialized with a wave field with the right analytic form and for wavenumbers and frequencies that satisfy the dispersion relation, the model must maintain this wave field forever.

For a more complicated balance, consider modons [33]. Modons are propagating solitary-wave like solutions that translate with constant velocity, while preserving their shape. Their dynamics are strongly non-linear. For a test, one might assert that the anomalies in a modon's shape are minimal in a frame of reference translating with the expected translation speed. Though complex, specific analytic solutions are available (for example see Boyd [3], their Section 16.14; eqns 16.54–16.56). Again, if initialized properly the model must maintain and propagate the modon without dispersion.

2.1.4 Using invariants for validity testing

Hughes [17] mentions invariants in the context of testing the code that generates input for property tests. This idea is particularly applicable to the balanced flow invariant. For example, we can test our specification of geostrophic balance using additional property tests. First, geostrophic flows are non-divergent so that $u_x + v_y = 0$. Second, the geostrophic velocity vector (u, v) is at a right angle to pressure gradient vector $(\partial p / \partial x, \partial p / \partial y)$. The latter property is evocatively taught to students: flow around a low (high) pressure is counterclockwise (clockwise) in the Northern Hemisphere; the directions are reversed in the Southern Hemisphere. The symmetry properties (Section 2.1.2) are potentially quite useful in this context too, since they apply to arbitrary flows.

2.2 What is the postcondition?

To quote Hughes [17]: “Postconditions relate return values to arguments of a single call. ... We can finesse this problem using a very powerful and general idea, that of constructing a test case whose outcome is easy to predict.”

Example: If we insert the minimum representable integer k into a list of integers and sort that list, k must be the first element of the sorted list.

What can we insert in to the ocean? Energy, through either heat (potential energy) or momentum (kinetic energy). The ocean *adjusts* to imparted energy through transients (usually radiating waves) and eventually evolves to a balanced state, usually geostrophic balance, assuming the system is not unstable. Nominally a property test might assert that stable systems will *eventually* reach a balanced state. Doing so would require very long integrations, greatly limiting the usefulness of such a test. In some rare cases, exact analytic solutions with transients may be calculated [13, their Section 7.3; pg. 196]. For a more general approach, I propose crafting anomalies that result in transients of known properties.

2.2.1 Resonant frequencies resonate

The balanced wave idea can be extended to include atmospheric forcing of an ocean basin. For simple basin configurations, we can analytically determine the dispersion relation of a whole host of balanced waves. These “free modes” of the system can be freely excited by atmospheric forcing of the same wavenumber and frequency — called “resonance”. The “postcondition” of inserting energy at resonant frequencies is energetic motion. At these frequencies, energy is very efficiently imparted in to the ocean. Conversely, no amount of energy input will excite energetic motions at frequencies or wavenumbers that are not on a dispersion relation. The “postcondition” of inserting energy at non-resonant frequencies is damped motion that dies quickly. Both are testable properties.

2.2.2 Symmetry yields asymmetry

A non-zero β parameter introduces a fundamental east-west asymmetry. In the mid-latitudes (poleward of $\approx 20^\circ$), the westward propagating Rossby wave modes are faster and have larger wavelengths. Eastward propagating modes are smaller wavelength and are dissipated by friction faster. This fundamental east-west asymmetry of response yields a useful property. Inserting energy at the western boundary of a square basin, should yield a response of eastward-propagating waves with a much shorter longitudinal length scale than inserting energy at the eastern boundary. The latter should yield a response of westward propagating waves of comparably larger length scale.

2.3 Metamorphic relations

To quote Hughes [17]: “Related calls return related results.”

Example: sorting a sorted list must yield the same list, or sorting any permutation of a list must yield the same result.

Metamorphic property relates the outcomes of two related experiments. In essence, this is the goal of GFD — to tease out parametric dependencies or functional relationships, or identifying controlling parameters [36]. In simple systems, a handful of parameters (e.g. stratification N , β , domain size) determine fundamental scales of balanced flows, and the predicted relationships can be easily tested. For example, a doubling of β should double the phase speed of a Rossby wave, or halve the time taken to receive a signal from the eastern boundary. Similar metamorphic relations can be constructed for all the proposed property tests earlier.

Metamorphic relations might offer a way to handle complex transient fields. Consider a rare case where the exact time-dependent analytic solution is known. Gill [13, their Section 7.3] discuss the analytic solution to the adjustment of a bump in sea surface height for a rotating constant density fluid. The solution has two parameters $P \equiv \eta_0 \sqrt{g/H}$ and “deformation radius” $L_D \equiv \sqrt{gH}/f$, where g is the acceleration due to gravity, H is the water depth, f is the rotation rate, and η_0 is the initial height of the bump. Again metamorphic tests can be designed by co-varying g, H, f, η_0 so that P and/or L_D are kept constant. Though not too useful when the analytic solution is known, we may apply the metamorphic relation when the analytics solution is unknown but the controlling parameters are known (e.g. the deformation radius L_D). This proposal will need extensive study but seems plausible. As an example, for the asymmetric response test (Section 2.2.2) we might insert wind forcing in the center of a square basin, and integrate for $\beta > 0$; and for $\beta < 0$. Presumably these solutions should be anti-symmetric, and easily comparable.

Finally, like all branches of fluid physics, GFD relies greatly on the concept of “dynamical similarity” — the idea that the core physics of a problem is determined by a handful of non-dimensional parameters. For example, there is a flow regime where the response of flow of speed U over sinusoidal topography of wavelength k on the β -plane is governed by the non-dimensional parameter $\lambda = Uk/\beta$ [13, pg. 525]. If U , k , and β were all individually rescaled so that λ remained constant, the solution for all these possible combinations of parameters is the same after appropriate rescaling. This statement is a metamorphic test derived from physics.

2.4 Inductive properties

Example: All subsets of a sorted list are sorted.

It is not apparent to me how the idea of induction might apply to a fluids system.

2.5 Model-based properties

Model-based tests compare the system under test to a much simpler reference implementation.

The previous sections are suffused with model-based properties given that our test ideas are derived from theoretical studies of a simplified model of the full equation set. In some cases, these simplifications can be relaxed through the use of simpler numerical models. For example, the dispersion relations for waves in systems with complex bottom topography and density variations can be hard to determine analytically, but possible to derive numerically [4]. Such numerically determined dispersion relations may be used to construct property tests for the more complex ocean model. Similarly, modon and other solitary wave solutions are hard to construct analytically but may be determined from simpler numerical models [9].

3 Discussion

Fluid physics naturally yields a large number of dynamical properties that may be recast in a form suitable for generative property-based testing. In a way these proposed tests are integration tests that exercise the ability of a model to reproduce expected physics of a simplified system. There is no new physics here. The core innovation here is to recast known ocean physics in the form of property tests, guided by the expertise of the computer science community. That many of the proposed testing ideas are already implemented as example problems or regression test cases in many ocean modeling systems, suggests that this approach may be profitable.

It is quite apparent that the core physics challenge here is the handling of transient motions. Either one uses flows that are balanced and hence excite no transient motions (Section 2.1.3), or one specifies initial conditions that excite transient fields of known properties (Section 2.2). Metamorphic relations, as expressed through the dynamical similarity of fluid flows, may also offer a way out of the difficulties posed by transients (Section 2.3), though more exploration of that idea is necessary.

It is possible to specify property tests at the level of individual terms in the equations, or for specific physical processes. As an example, consider the submodule that handles advecting (or moving) a “tracer variable” like dye, temperature, or salinity. In a domain that is both periodic in x and y , there is a background velocity (u, v) for which a patch of dye will return to its starting location after N time steps. The challenge is whether the codebase is modular enough to allow submodules to be tested individually.

Two major unaddressed challenges here are:

1. Can we design appropriate shrinking strategies? Shrinking is the idea that generated test cases can be shrunk to yield simpler and more minimal test cases. Shrinkers should be designed such that the simplifying physics assumptions underlying the test are not broken.
2. Can we limit the computational expense of the proposed tests to make them feasible?

A concrete implementation of the proposed ideas will be necessary to answer these two questions, and to provide guidance on which property tests are most useful at finding bugs.

I end with a final quote from Hughes [17]: “every bug is found by at least one postcondition, metamorphic property, and model-based property.” Luckily we have many ideas that lie in those three categories!

References

- [1] Alper Altuntas & John Baugh (2018): *Hybrid Theorem Proving as a Lightweight Method for Verifying Numerical Software*. In: *2018 IEEE/ACM 2nd International Workshop on Software Correctness for HPC Applications (Correctness)*, IEEE, Dallas, TX, USA, pp. 1–8, doi:10.1109/Correctness.2018.00005. Available at <https://ieeexplore.ieee.org/document/8638360/>.
- [2] William Blumen (1972): *Geostrophic Adjustment*. *Reviews of Geophysics* 10(2), pp. 485–528, doi:10.1029/RG010i002p00485. Available at <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/RG010i002p00485>.
- [3] John P. Boyd (2018): *Dynamics of the Equatorial Ocean*. Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/978-3-662-55476-0_9. Available at http://link.springer.com/10.1007/978-3-662-55476-0_9.
- [4] Kenneth H. Brink (2018): *Coastal-Trapped Waves with Stratification, Topography, Mean Flow and Bottom Friction with Complex Frequency in Matlab*. Available at <https://hdl.handle.net/1912/10525>.
- [5] Tsong Yueh Chen, Joshua Wk Ho, Huai Liu & Xiaoyuan Xie (2009): *An Innovative Approach for Testing Bioinformatics Programs Using Metamorphic Testing*. *BMC Bioinformatics* 10(1), p. 24, doi:10.1186/1471-2105-10-24. Available at <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-10-24>.
- [6] D. A. Cherian, D. B. Whitt, R. M. Holmes, R.-C. Lien, S. D. Bachman & W. G. Large (2021): *Off-Equatorial Deep-Cycle Turbulence Forced by Tropical Instability Waves in the Equatorial Pacific*. *Journal of Physical Oceanography* 51(5), pp. 1575–1593, doi:10.1175/JPO-D-20-0229.1. Available at <https://journals.ametsoc.org/view/journals/phoc/51/5/JPO-D-20-0229.1.xml>.
- [7] Koen Claessen & John Hughes (2000): *QuickCheck: A Lightweight Tool for Random Testing of Haskell Programs*. *ACM SIGPLAN Notices* 35(9), pp. 268–279, doi:10.1145/357766.351266. Available at <https://dl.acm.org/doi/10.1145/357766.351266>.
- [8] CROCO developers: *15. Review of Test Cases — Croco Doc 2.0.1 Documentation*. Available at https://croco-ocean.gitlabpages.inria.fr/croco_doc/model/model.test_cases.html.
- [9] Matthew N. Crowe & Edward R. Johnson (2024): *Modon Solutions in an N-layer Quasi-Geostrophic Model*. *Journal of Fluid Mechanics* 994, p. R1, doi:10.1017/jfm.2024.619. arXiv:2404.07718.
- [10] Olivier Darrigol (2005): *Worlds of Flow: A History of Hydrodynamics from the Bernoullis to Prandtl*. Oxford University Press Oxford, doi:10.1093/oso/9780198568438.001.0001. Available at <https://academic.oup.com/book/54747>.
- [11] Charles L Fefferman (2006): *Existence and Smoothness of the Navier-Stokes Equation*. In: *The Millennium Prize Problems*, Clay Mathematics Institute, pp. 57–70.
- [12] Joseph Galewsky, Richard K. Scott & Lorenzo M. Polvani (2004): *An Initial-Value Problem for Testing Numerical Models of the Global Shallow-Water Equations*. *Tellus A: Dynamic Meteorology and Oceanography* 56(5), p. 429, doi:10.3402/tellusa.v56i5.14436. Available at <https://a.tellusjournals.se/article/10.3402/tellusa.v56i5.14436/>.
- [13] Adrian E. Gill (1982): *Atmosphere-Ocean Dynamics*. International Geophysics Series, Academic Press, New York.
- [14] Isaac M. Held & Max J. Suarez (1994): *A Proposal for the Intercomparison of the Dynamical Cores of Atmospheric General Circulation Models*. *Bulletin of the American Meteorological Society* 75(10), pp. 1825–1830, doi:10.1175/1520-0477(1994)075<1825:APFTIO>2.0.CO;2. Available at [http://journals.ametsoc.org/doi/10.1175/1520-0477\(1994\)075<1825:APFTIO>2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0477(1994)075<1825:APFTIO>2.0.CO;2).
- [15] Haosheng Huang, Changsheng Chen, Geoffrey W. Cowles, Clinton D. Winant, Robert C. Beardsley, Kate S. Hedstrom & Dale B. Haidvogel (2008): *FVCOM Validation Experiments: Comparisons with ROMS for Three Idealized Barotropic Test Problems*. *Journal of Geophysical Research: Oceans* 113(C7),

- p. 2007JC004557, doi:10.1029/2007JC004557. Available at <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2007JC004557>.
- [16] Chris W. Hughes, Joanne Williams, Angela Hibbert, Carmen Boening & James Oram (2016): *A Rossby Whistle: A Resonant Basin Mode Observed in the Caribbean Sea*. *Geophysical Research Letters* 43(13), pp. 7036–7043, doi:10.1002/2016GL069573. Available at <https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2016GL069573>.
 - [17] John Hughes (2020): *How to Specify It!: A Guide to Writing Properties of Pure Functions*. In William J. Bowman & Ronald Garcia, editors: *Trends in Functional Programming*, 12053, Springer International Publishing, Cham, pp. 58–83, doi:10.1007/978-3-030-47147-7_4. Available at http://link.springer.com/10.1007/978-3-030-47147-7_4.
 - [18] John Hughes (2020): *John Hughes - Keynote: How to Specify It! Lambda Days 2020*. Available at <https://www.youtube.com/watch?v=G0NU0st-53U>.
 - [19] W. G. Large, J. C. McWilliams & S. C. Doney (1994): *Oceanic Vertical Mixing: A Review and a Model with a Nonlocal Boundary Layer Parameterization*. *Reviews of Geophysics* 32(4), pp. 363–403, doi:10.1029/94RG01872. Available at <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/94RG01872>.
 - [20] David Lawrence: *Community Earth System Model 3 (CESM3): Plans, Progress, Timelines*. Available at <https://www.cesm.ucar.edu/news/community-earth-system-model-3-cesm3-plans-progress-timelines>.
 - [21] David MacIver, Zac Hatfield-Dodds & Many Contributors (2019): *Hypothesis: A New Approach to Property-Based Testing*. *Journal of Open Source Software* 4(43), p. 1891, doi:10.21105/joss.01891. Available at <https://joss.theoj.org/papers/10.21105/joss.01891>.
 - [22] Andrew J. Majda & Andrea L. Bertozzi (2001): *Vorticity and Incompressible Flow*, 1 edition. Cambridge University Press, doi:10.1017/CBO9780511613203. Available at <https://www.cambridge.org/core/product/identifier/9780511613203/type/book>.
 - [23] M.E. McIntyre (2003): *Balanced Flow*. In: *Encyclopedia of Atmospheric Sciences*, Elsevier, pp. 680–685, doi:10.1016/B0-12-227090-8/00140-8. Available at <https://linkinghub.elsevier.com/retrieve/pii/B0122270908001408>.
 - [24] M.E. McIntyre (2015): *Potential Vorticity*. In: *Encyclopedia of Atmospheric Sciences*, Elsevier, pp. 375–383, doi:10.1016/B978-0-12-382225-3.00140-7. Available at <https://linkinghub.elsevier.com/retrieve/pii/B9780123822253001407>.
 - [25] MITgcm Developers: *MITgcm Testing Protocols*. Available at <https://mitgcm.readthedocs.io/en/latest/contributing/contributing.html#mitgcm-code-testing-protocols>.
 - [26] MOM6 developers: *MOM6 Test Suite*. Available at <https://github.com/mom-ocean/MOM6/tree/main/.testing>.
 - [27] MOM6 developers: *Novel Tests — MOM6 0.2a3 Documentation*. Available at <https://mom6.readthedocs.io/en/main/api/generated/pages/Testing.html#novel-tests>.
 - [28] Joseph Pedlosky (1987): *Geophysical Fluid Dynamics*, 2. ed., nachdr edition. Springer Study Edition, Springer, New York.
 - [29] L. M. Polvani, R. K. Scott & S. J. Thomas (2004): *Numerically Converged Solutions of the Global Primitive Equations for Testing the Dynamical Core of Atmospheric GCMs*. *Monthly Weather Review* 132(11), pp. 2539–2552, doi:10.1175/MWR2788.1. Available at <http://journals.ametsoc.org/doi/10.1175/MWR2788.1>.
 - [30] J.F. Price (2006): *A Coriolis Tutorial*. Available at https://www2.whoi.edu/staff/jprice/wp-content/uploads/sites/199/2019/01/aCt_2003.pdf.
 - [31] ROMS developers: *Test Cases - WikiROMS*. Available at https://www.myroms.org/wiki/Test_Cases.

- [32] Simone Silvestri, Gregory Wagner & R. Ferrari (2024): *The World's Fastest Ocean Model*. Available at <https://clima.caltech.edu/2024/04/15/the-worlds-fastest-ocean-model/>.
- [33] Melvin E Stern (1975): *Minimal Properties of Planetary Eddies*. *Journal of Marine Research* 33(1), pp. 1–13.
- [34] Philip D. Thompson (1988): *Haurwitz Solutions of the Nonlinear Shallow-Water Equations for Small Froude Number*. *Meteorology and Atmospheric Physics* 38(1-2), pp. 89–94, doi:10.1007/BF01029950. Available at <http://link.springer.com/10.1007/BF01029950>.
- [35] Geoffrey K. Vallis (2016): *Geophysical Fluid Dynamics: Whence, Whither and Why?* *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 472(2192), p. 20160140, doi:10.1098/rspa.2016.0140. Available at <https://royalsocietypublishing.org/doi/10.1098/rspa.2016.0140>.
- [36] Geoffrey K. Vallis (2017): *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*, 2 edition. Cambridge University Press, doi:10.1017/9781107588417. Available at <https://www.cambridge.org/core/product/identifier/9781107588417/type/book>.
- [37] Hillel Wayne (2019): *Finding Property Tests*. Available at <https://www.hillelwayne.com/post/contract-examples/>.
- [38] David L. Williamson, John B. Drake, James J. Hack, Rüdiger Jakob & Paul N. Swarztrauber (1992): *A Standard Test Set for Numerical Approximations to the Shallow Water Equations in Spherical Geometry*. *Journal of Computational Physics* 102(1), pp. 211–224, doi:10.1016/S0021-9991(05)80016-6. Available at <https://linkinghub.elsevier.com/retrieve/pii/S0021999105800166>.
- [39] Scott Wlaschin (2014): *The 'Property Based Testing' Series | F# for Fun and Profit*. Available at <https://fsharpforfunandprofit.com/series/property-based-testing/>.
- [40] Xarray developers: *Xarray Property and Stateful Tests*. Available at <https://github.com/pydata/xarray/tree/main/properties>.
- [41] Zarr Python developers: *Zarr Python Property and Stateful Testing*. Available at <https://github.com/zarr-developers/zarr-python/tree/main/src/zarr/testing>.