Next Generation Ocean Dynamical Core Roadmap Project: Summary and Recommendations

Jason Holt, Adrian New, Hedong Liu, Andrew Coward (NOC)
Stephen Pickles, Mike Ashworth (STFC)

1. Introduction

This document provides a summary of the final report of the Next Generation Ocean Dynamical Core Roadmap Project and recommendations for the way forward: essentially it provides the 'road map'. To be read alongside the final report, it describes two complementary ways forward for ocean modelling in the UK. First is the incremental evolution of the NEMO model, and second a new modelling initiative drawing on the GungHo project; we make the assumption that the current NEMO code base has a finite competitive life time, and over the course of this review period will 'lose its edge' and become increasing inefficient and problematic to use. The key question is then whether this existing code base can be re-factored for future computer architectures or whether a new approach is needed. In this consideration, it is vital to make the distinction between the NEMO code and the NEMO consortium; the latter is seen as a crucial element in UK ocean modelling capability throughout the review period.

The Ocean Road Map project has identified two concurrent pathways for ocean model development in the UK:

1. **Develop NEMO for global, shelf sea and 'global coastal ocean' applications.**
2. **Develop an ocean model within the GungHo framework.**

This document largely focuses on the justification for (2), as being the new direction. Further details on (1) for the short- medium term can be found in section 2 and 3 of the final report. .

2. The computational landscape in the UK over the 5-20 years

Marine modelling is closely linked to scientific High Performance Computing, and this is turn to the trends in this technology, largely driven by external factors. The research council HPC facility provides a good guide to future computing capability in the UK. There are other facilities available e.g. the Met Office computer, the joint Met Office/NERC computer, capability computing services such as HARTREE, and local clusters. Here we assume these will grow at a comparable rate to the UKRC service.

From HPCx, through the four phases of Hector to the expected size to ARCHER the peak performance of the UKRC HPC facility has increased exponentially over the past ~8 years. Given this trend has somewhat flattened off since the rapid increase between HPCx and Hector Phase 2a, the conservative estimate is to extrapolate the trend from Phase 2a to Archer. This gives a peak performance of ~50 times HECTOR Phase 3 by 2019 (100Pflop/s) and ~1000 times Phase 3 by 2023(5Eflop/s). This closely follows Moore's Law / TOP500 trends, and predicts the UK maintains a performance about a factor of ten lower than the US at any one time (or lags by 3-4 years). There are of course many unknowns in this prediction including the UK Governments continued commitment to HPC, and the share of the resource the UK marine community may receive, but there are presently no indications that these will falter.

The current expectation is that this increase in computer power will be achieved through a substantial increase in core count per physical chip and memory amount/band width will not keep up with this increase; hence memory per core is expected to steadily decrease. For example, memory per core has decreased from 3GB to 1GB over the life of HECTOR. Of course there may be radical changes in computer architectures that overcome the present power barrier[1], but we cannot predict these. Hence, it is prudent to assume the current trends will persist. **The present approach of one MPI task per core will soon become impractical as there will not be the memory available to support an optimal fraction of the model grid on each core on a chip. Hence, alternative approaches will be needed and mixed MPI, openMP is a very attractive option.** When this break-point will occur with NEMO is difficult to predict as this model is primarily limited by memory bandwidth rather than amount. However, HECTOR Phase 3 may be close to this limit (e.g. NEMO AMM will run optimally on up to 320 cores, but fails when run on 160 or fewer cores, so a x2 increase in cores per node without increasing memory may cause problems). Hence, we can expect difficulties over the course of ARCHER and certainly on the follow on computer (i.e. by 2019). The implications of doing nothing will be that to run a given NEMO configuration, we will have to increasingly underutilise the available cores to gain memory bandwidth. This means **NEMO will become increasingly expensive to run over time and, if clock speeds decrease, our capability to do science with it will diminish rather than increase**.

3. User drivers for future ocean model development

Here we consider the user drivers for an increasing ocean modelling capability, which will require us to meet the computational challenges identified above.

3.1 Incremental model evolution

The drive for finer resolution ocean models is tensioned against the requirements for longer simulation, more ensembles and more complex representation of different components of the system (e.g. biogeochemistry, ecosystems etc). From a physical oceanographic point of view, the need to resolve motions at the first baroclinic Rossby Radius and key geographic features is paramount. There is also a need to avoid grey areas of resolution where the qualitative nature of the model (e.g. eddy resolving or not) is ambiguous; unfortunately the current workhorse global model (ORCA025) sits in such an area. Given that the success of other modelling activities is predicated on the physics there is a very strong case to move toward eddy resolving models (e.g. $1/12^o$) as the 'workhorse' resolution over the period under consideration (5-15 years). Similarly, in coastal, shelf sea models finer resolution (e.g. $1/60^o$) gives a clear benefit in terms of resolving coastline/topography and local detail (e.g.

[1] Power to individual chips is limited to about 100W, so that clock speeds have now peaked at around 2-3 GHz

parameterising wind farms), and dynamical features such as frontal jets, upwelling, and tidal excursions.

The question then is: should we plan to go beyond this with the current generation of structured grid models, with an ever refining ‘high resolution everywhere’ approach? At the global scale, there is a case to move beyond $1/12^{o}$ (to say $1/36^{o}$) to ensure the mesoscale eddies, upwelling systems, basic continental shelf features, and pinch points are well resolved. However, the case for going further is less clear. The resolutions we are considering are far from convergence, so subgrid scale parameterisations (other than LES type) of the sub-mesoscale will still be required even as we start to resolve this scale (i.e. we move into the next grey area). Also, questions arise of whether data handling infrastructure will be able to keep-up. Hence, at $1/36^{o}$ and beyond, the case for multi-scale modelling becomes increasingly strong, so that resolution can be targeted where it is most needed according to coast line and topography, and dynamics (e.g. where the Rossby radius is small such as in the Arctic). Two-way nesting provides an option for some aspects of multi-scale modelling where dynamic (e.g. the Arctic) or user driven (e.g. European seas) need is obvious. It does, however, substantially lack flexibility beyond this (e.g. to refine according to global criteria). **While the move to finer resolution and eventually multi-scale approaches is crucial to maintain the international competiveness of ocean modelling in the UK, the user requirement for efficient coarser resolution structured grid models will endure throughout the review period. These must be effectively accommodated within any solution.**

Another important question is: at what point will the current code base (NEMO) become inefficient/ineffective on evolving computer architectures, given the challenges noted above. This depends on how much effort is spent on incrementally developing this code base to meet these challenges. However, there will come a point where the recoding to meet these is so intensive and extensive that the activity really constitutes a complete rewrite of the code. Moreover, incrementally optimising the code, without radically restructuring it (for example to separate science and computer-science layers, as in GungHo) will seriously hamper the ability of the community to maintain the code, its usefulness and its ‘developability’. Hence, through the use of mixed MPI and openMP, and advanced IO management approaches we see no serious scalability bottlenecks at least in the short to medium term, so long as this development is adequately resourced in a timely fashion. However, **the serious challenge lies in meeting the optimisation requirements and maintaining a code base usable by its current broad community. We do not believe this is achievable with the current NEMO code after about 2019.** There are essentially two options on the table to address this:

A. A science neutral rewrite of the NEMO code

B. Moving to a new modelling approach, as described below.

Option (A) could be conducted either using the GungHo approach (below) or an alternative coming out of the NEMO consortium.

The UK also makes use of several other model systems from coastal to open ocean, including structured and unstructured grid approaches. These will also face similar computational challenges and the originators largely take responsibility for their future. However, the detailed pathways and levels of support are not necessarily clear. At this stage we do not propose using NC funding to support the development or optimisation of any of these models; of course funding from others UK sources may be forthcoming. A particular issue lies in the shelf-coastal modelling community. This group is increasingly using the FVCOM model and there may be a risk to this community if this model does not maintain its international competitiveness; however, it is currently too early to judge this. There is already a strong UK working group on FVCOM modelling (at NOC, PML, SAMS). We recommend that this group be supported and that **a review of near coastal modelling needs is conducted in 3-5 years time (aligned with appropriate project cycles)** to assess currently available model options**.**

3.2 New directions in ocean modelling

Beyond the incremental refinement of existing models briefly considered above, is the question of 'what are the important new areas of activity that would meet the up and coming users demands?' These are described as the 'drivers for change' in the Final Report but to summarise: **DFC1**: Linking Coastal, Shelf and Open Ocean Models; **DFC2**: Improved Mixing; **DFC3**: New Computer Architectures. Of these it is DFC1 and the general flexibility of multi-scale modelling (noted above) that is the primary driver toward unstructured grid modelling approaches. The unstructured grid approach is obviously applicable to the near coastal and shelf scale modelling. However it has a wide range of benefits for open-ocean and ocean-shelf modelling, particularly regarding topography (e.g. areas of restricted exchange) and regions of interest. However, this must be tensioned against the 'high resolution everywhere' option. **At a global scale an unstructured grid approach is unlikely to be beneficial at resolutions coarser than ~$1/12^{o}$** (although the flexibility would still aid regions of interest).

The enduring need for lower resolution models (global and regional) must be borne in mind, and any new approach should accommodate these at a computational efficiency at least comparable to existing models. There may be an element of 'running to stand still' with lower resolution models, if computer clock speeds do decline and they must be run on grossly under-populated compute nodes (to gain memory) if there is simply not the concurrency in the problem at hand. It is for these that option (A) above becomes attractive.

An immediate application for unstructured grid modelling would be a UK or European shelf-coast model. This could, for example, cover the Northwest European continental shelf refined to particular areas of interest, activity or jurisdiction, according to the important physical, ecological or operational scales; e.g. refined around the UK EEZ. The question remains whether every bay and estuary in the UK or NW European should be included? A practical approach would be to include all the major estuaries and have alongside this a rapid reconfiguration capability that would allow refinement to other areas as need arises.

3.3 Summary of user drivers

It is helpful to consider a range of possible model configurations that would meet the user requirements over the period and when these become practical as 'routine' physical models (i.e. ocean components of IPCC input and operational models) or routine for ensemble runs and with embedded ecosystems. Costs can be estimated on the basis of number of grid cells, timestep and a penalty for using unstructured meshes (taken to be 5) and for ecosystems (taken to be 8, as an upper bound of their cost). Dates can then be assigned to each grid development using the projected exponential grow of the UKRC computer facility, accepting that there are many caveats to this. From Table 1, it is clear that treatment of time stepping is the major consideration as resolutions are refined, particularly in the near coastal zone. Beyond 1/12^{o} global and 1/60^{o} shelf sea, unstructured grid approaches become increasingly attractive. This is particularly the case given these estimates suggest it would be three computer refreshes after HECTOR before a 1/36^{o} global model would be considered routine; a very challenging proposition for NEMO without a complete rewrite. **Hence, by the time the current generation of advanced models has becomes routine, computer architectures will have developed to the extent that the next generation of advanced models will require a radical change before they too can become routine.**

Table 1: Possible model grids

Grid	S/US	Vertical	Size	Cost	Cost	When routine physics model	When routine ensemble/BGC model
			(k cells)	(time step)	(no time step)		
				cf ORCA025			
Global Scale							
¼	S	75 Z	904	1	1	2011	2014
1/12	S	75 ALE	8150	27	9	2016	2020
1/36	S	100 ALE	73350	973.7	108.2	2022	2026
1/4+1/12 multiscale	US	75 ALE	2802	46.5	15.5	2017	2021
1/12+1/36 multiscale	US	100 ALE	8700	577.4	64.2	2021	2025
Basin Scale							
1/12 (NA)	S	75 s-Z	1080	3.6	1.2	2013	2016
1/12+1/60 NWS	US	75 ALE	2856	189.6	15.8	2020	2023
Shelf scale (NWS)							
1/12	S	50 s	111	0.2	0.1	2008	2012
1/60	S	50 s	1776	15.7	1.3	2015	2019
1/120	S	75 s	7104	125.7	5.2	2019	2022

Shelf-coast							
1/160+200m UK coast	US	50 s	3448	1346.2	12.7	2023	2026
1/160+200m NW Europe coast	US	50 s	5030	1963.8	18.5	2023	2027
1/160+200m UK + 50m Estuary x N	US	50 s	3448+N*400m	12358.4	27.5	2026	2030

S= structured, US = unstructured

4. New approaches for Ocean Modelling in UK and Europe

The need for a new approach for ocean modelling is clear if we are to both maintain competitiveness on future computer architectures, and exploit the potential of multi-scale modelling. **To this end we advocate aligning an ocean model development effort with the GungHo project, and propose the resulting model forms the new dynamical core for use by the NEMO consortium**. GungHo (see section 4.1) is a large UK effort involving NERC, the Met Office and the UK Academic community, to build a new dynamical atmosphere core for the UK, and represents an immense opportunity for ocean modelling, that the whole UK community can engage with. It also presents an opportunity to move the NEMO consortium to a second phase, building on the strengths of the consortium, but releasing it from a code base that may become increasingly outdated. It is quite possible that the consortium may choose a different path for advanced ocean model development or adopts a 'wait and see' stance, but we should put this option on the table and actively engage with those partners who are interested. **If the other consortium members choose a different route then the UK partners would need to consider very carefully before 'going it alone' with a GungHo option.**

4.1 The GungHo Project

The GungHo project proposes to develop code in separate components: a computational science or 'driver' component, which is likely to be derived from freely available generic software tools, and the model components which solve the model equations. The philosophy behind this structure is that the scientific equations (which are coded as algorithms in an Algorithmic layer and computed using reusable local kernel operators in a Kernel layer) are expected to be coded primarily by ocean/atmospheric scientists, and are kept separate from the computational infrastructure, which is the domain of the computational scientists. There is a separation between the Parallel System component, where parallelization, communications and computational tasks are coded, and the Algorithmic and kernel components. The algorithmic layer is primarily where the model specific equations will exist, although it is expected that some aspects of the kernels may also have to be updated with changes to the scientific equations.

The separation of the computational science layer from the natural science layers is the crucial point: it isolates the natural scientist as far as possible from the complexities needed to ensure efficient scalability. The expectation is that this computational science layer will provide a set of tools and approaches, many of which will be easy to adapt for ocean as well as atmospheric modelling, while a number of the kernel operators will also be applicable in an ocean model. The degree of flexibility available for the design of the ocean model has yet to be established (i.e. how far it would be able to deviate from the atmospheric approach). The framework is likely to be able to accommodate differing element shapes and types, and vertical grid approaches, but differing solution strategies (e.g. choosing a Finite Volume rather than a Finite Element approach) are likely to be more problematic and require more ocean specific development effort, or indeed may not be practical.

4.2 Properties of an ocean model

Before embarking on any ocean model development effort it is necessary to identify the desirable properties of an ocean model, priorities these, and then explore whether they can be achieved with different approaches.

A list of desirable properties is:

1. Good discrete dissipation properties: i.e. minimizes numerical diffusion
    a. Permits features commensurate to resolution
    b. Deep water mass properties are maintained over decadal timescales
    c. Can accommodate theoretically/empirically sound subgrid scale models
2. Conservation of mass, tracer, energy, momentum, PV, and perhaps enstrophy
3. Good discrete dispersion properties
    a. Numerical modes are controllable without breaking other requirements (particularly 1)
4. Computationally efficient
    a. For large, high resolution models and smaller models needed for ensemble and long simulations
    b. Increase in computational cost for similar accuracy is much less than increase in computer power over the development time
5. Fits coastline at least as well as quadrilaterals, ideally as well as triangles
6. Flexible vertical coordinates and methods (including ALE)
7. Multiscale horizontal mesh capability and so geometric flexibility
8. Accurate in realistic and idealized test cases
9. Portable, adaptable and easy to implement
    a. Can be effectively used by a small community and small groups (e.g. 1PI, 1 post doc, 1 phd student)
    b. Has a well supported infrastructure
    c. Can accommodate future developments

Against these properties the range of possible solution approaches (Finite Element, Difference, Volume) and grid arrangement need to be tabulated to provide a ‘Properties-Approaches’ matrix. It is expected that no combination of solution approach and grid arrangement will meet all of these criteria, and moreover it is not practical to objectively assess all of these in realistic tests (indeed in several cases an approach for testing is not well established). Hence we must fall back on expert judgment, past lessons and compromise.

4.3 Engaging with the GungHo Process

The recommendation here is to develop a separate GungHo Ocean (G-Ocean) science layer that shares, with the atmospheric code, the computational science layer and where possible solution approaches, but has the flexibility to develop in a somewhat different direction from that of the atmosphere. For example, the GungHo project may well decide on a FE model on quadrilateral elements. This would be a rather perverse choice for an ocean model as it does not accommodate improved coastline matching or a multi-scale capability, but comes at the expense/effort of the FE approach.

We propose a three stage approach in engaging with the GungHo project:

Stage 1: Planning, prototyping and consortium building (Commencing early-mid 2013) 2013-2014

Activity A: (Planning) A G-Ocean steering committee will be convened in 2013 with the objective of exploring the Properties-Approaches matrix. This will be a group of O(10) PIs and experts largely drawn from outside the GungHo community, but with some cross-over representation .

Activity: B: (Prototyping) a group of modelling practitioners will be gathered to explore test cases and help to inform (A). Hopefully this will be able to use the same tools as GungHo and will be able to give modellers early sight/experience of how the code is developing.

Activity C: (International consortium building). We will open discussions with the partners in the NEMO consortium as to the viability of developing G-ocean as the follow on to OPA as the core code in NEMO, i.e to become NEMO2. Mechanisms include: Presenting Ocean Road Map documents to the NEMO steering committee as the UK's vision for the evolution of the NEMO consortium; lobbying to get appropriate call texts into Horizons 2020; informal discussion with key players.

Gateway 1: Will this approach perform better than NEMO (OPA) given its current trajectory? Is funding in place for Stage 2? 2014

Stage 2: Construction of the basic dynamic model.

A consortium (UK and EU if possible) would be assembled to construct a global and regional 3D model. 2014-2017

Gateway 2: Commitment for this to become the next strategic ocean model; formal engagement with the NEMO apparatus begins here. 2017

Stage 3: Development of a fully-fledged ocean model. Expand the consortium to include sea-ice, waves, biogeochemistry, ecosystems and data assimilation. 2017-2021