Government Office for Science

Review of climate science advice to Government: supporting paper on underpinning high performance computing / supercomputing requirements

1. Why is enhanced supercomputing capability needed?

There is a need to develop and run model resolutions which can resolve key processes, i.e. at the sort of resolution used in current numerical weather prediction (of the order of 50km in the atmosphere). A longer term “holy grail” is to work towards modelling scales which resolve features such as deep convective cloud systems and ocean eddies (of the order of 1km in both the atmosphere and ocean).

Progress in addressing key questions will require both scientific and technological advances. The ability to run climate models to give the best advice is limited by:

- *Scientific development*, including the development of more complete Earth system processes and the reformulation of physical processes to work properly in much higher resolutions models;

- *Model performance*, best measured by time to solution. The model needs to run quickly enough to give results in a useful timescale. Model performance is partly limited by supercomputer performance. However, because processor speeds are not improving, the dominating factor on model performance will be increasing model scalability;

- *Computing capacity*, which can be measured by the number of climate model runs that can be performed, and can be overcome by having a larger computing system.

At present, the most demanding limitations are on model performance and computing capacity. Scientific development is necessary, but is currently well ahead of performance and capacity. The pace of scientific development can also be improved to some extent by increasing performance and computing capacity.

The drive to high resolution models with full Earth system components will put an increasing strain on model performance. Calculations, based on known scalability of weather forecast models, show that the Unified Model, with 25km resolution atmosphere, cannot be run fast enough for climate change simulations no matter how much computing capacity we have, and it is unlikely that a 25km model could be run quick enough (e.g. at 2 model years per day) until 2018-2022 when intermediate improvements in the dynamics scheme could be implemented. Running a 1km global climate model with full Earth system components will not be possible until well after 2025. Climate models with enhanced Earth system components at 25km atmospheric resolution would require significantly more scalability developments such as a new dynamical core and increases in parallelism across components. Work on improved model scalability is underway and could be enhanced, but is unlikely to drastically reduce development time.

2. Trends in HPC

The following trends will shape and constrain the way in which supercomputing resource and models can best be used in the future:

2.1. Technology trends

Future supercomputing performance improvements will be derived from increases in system scale (rather than, as hitherto, improvements in single-processor performance). This will mean that current model codes, if not radically modified, will not run well on next-generation technology. Going from Tera-flop to Peta-flop systems and beyond means that the number of components (cores, interconnect, memory) within such systems will grow enormously. In the near future we will see clustered Multi-Core systems with core numbers in the range of one-hundred thousand to one million and more.

CPU manufacturers are continuing to improve manufacturing technology to utilise finer scale processes. This allows the number of transistors on a physical CPU to increase at the rate that Moore's Law[1] predicts. In the past this process shrinkage has allowed CPUs to run at higher clock frequencies (and thus be quicker) but fundamental physical constraints have now been reached. Heat generated by current leakage is now too great to continue the increase of clock frequency and all manufacturers are now releasing CPUs that actually run at lower speeds than those available a few years ago, containing increasing numbers of cores.

Performance increases are now largely achieved through ever-increasing numbers of commodity micro-processors. These undergo performance-enhancing logic refreshes on an annual basis, designed to be backwardly-compatible to minimise customer costs. It is therefore expected that the scientific programming environment will remain stable for the next 10-15 years. More-frequent performance-enhancing hardware refreshes than have been typical hitherto are now possible without recourse to model software re-writes. However there is a longer-term need for scalability issues to be addressed via a radical code re-design.

2.2. Scalability of software

Weather and climate model software is continuously optimised, with vendor support, to make best use of evolving technology and enable models to run within demanding timescales.

The numerical algorithms which form the basis of weather and climate models have structures which do not lend themselves to easy parallelism. A recent study commissioned by the Met Office concluded that commodity clusters would be unlikely to be cost-effective technology for running high-resolution climate models, but might be appropriate for running lower-resolution ensembles of 50 or more.

Current weather and climate codes (such as the Unified Model[2]) can be run on massively parallel systems but there is a natural limit to how many cores can be used before non-parallel parts of the code start to dominate. This scalability limit depends on the algorithms being used, model resolution and system performance and is increasingly dependant on the system's ability to move data from memory to processor and between processors.

---

[1] A prediction of the pattern of future performance improvements in computer hardware based on the observed long-term trend that integrated circuits double in performance every 18 months.

[2] The Met Office Unified Model is the numerical modelling system used widely for climate science in the UK and internationally. It is 'seamless' in that different configurations of the same model are used across all time and space scales. The different configurations are each designed to best represent the processes which have most influence on the timescale of interest.

To meet this challenge, the basic numerical algorithms used to solve the governing equations in climate and weather models will need to be redesigned to minimize the need for inter-processor communications, so next-generation technology can be harnessed to best effect to run high-resolution and other climate models. This work is currently estimated to take about ten years.

The Met Office and NERC plan to undertake further fundamental code re-design work, both in-house and in partnership with other institutes (see later section).

2.3. Power and cooling operating costs

Power and cooling requirements of future supercomputer systems could rise significantly thus posing major engineering and budgetary challenges.

Supercomputer manufacturers and data centres are focusing increasingly on developing energy efficient systems, for example through developing solid state disk technologies and CPUs that intelligently switch off chip circuits when not in use. Vendors are also experimenting with running systems at much higher temperatures in order to make free-to-air cooling more-efficient.

Currently, up to 50% of total data centre power demand is used for equipment cooling. Through a combination of better management of IT Hall layout and power supply, as well as greater take up of free-to-air cooling, it is expected that this will reduce to less than 10% within the next 5-10 years. Already, both DC power and ambient-air cooling energy efficient technologies[3] have been deployed across the Met Office, saving power costs and reducing emissions and these infrastructures can be upgraded to serve future more-powerful supercomputing facilities.

The Met Office and ECMWF are currently jointly appraising how ‘green’ technologies may best be better harnessed to provide data centre power and cooling.

2.4. Data archiving and retrieval needs

As supercomputers become ever more-powerful, and model resolutions increase, ever-larger volumes of data will be needed as input and output for storage and subsequent analysis (locally and remotely). The capacity and capability of data archiving, retrieval and sharing systems need significant attention; they will need to be increased alongside supercomputer performance.

2.5. Networks and Access to Resources (both HPC and data)

The proliferation of HPC coupled with increasing data input and output requirements has meant that the ability to move data around has become ever more important. The state of the art is now 10 Gbit/s access between significant HPC centres, and between them and major data archives, with roadmaps to even higher bandwidth in place at most institutions. On a European and global scale, major institutions are evaluating and/or putting in place “light paths” (dedicated network links) which will guarantee the required data movement bandwidth, and thus facilitate writing data to, and retrieving data from, remote data archives. With high-bandwidth networks, improved data analysis algorithms associated with fewer larger data archives are now preferred over large data stores with each machine, since ensemble statistics

[3] Funded via DECC’s *Low Carbon Technology* budget.

and model inter-comparisons can be carried out with fewer data movements. Effective networks are fundamental to meeting HMG needs, because of the multi-disciplinarity of the science and the consequent need to engage a wide community of expertise.

3. Climate simulation models and computing capability needed to generate the scientific evidence

3.1. The importance of model resolution

There is a need for climate simulations to be made ever more accurate at regional scales and the resulting predictions to be accompanied by better-informed levels of confidence. Higher resolution is required to better capture spatial detail, to better represent physical processes, and more accurately represent topography and land/sea boundaries.

Many of the damaging impacts of climate variability and climate change will be felt through climate extremes, especially changes in precipitation. A substantial increase in model resolution in the horizontal and vertical is an essential component of delivering improved information. Accompanying this must be a concerted effort to improve fundamental processes in climate models, particularly related to cloud and precipitation processes, such as organised tropical convection.

In parallel, it may be possible to use lower resolution models to address some policy questions, where these models have been designed to give robustly-traceable answers. However, it is not yet clear if this is possible and higher resolution models would be needed anyway to validate the technique.

There is a balance between giving the best possible advice on timescales that society needs, i.e. as soon as possible, versus the longer-term scientific aspiration to provide as accurate and reliable scientific advice as possible. Ideally, we would run global climate models, with full Earth system components at a resolution of about 1km in the atmosphere and ocean to resolve convective clouds and ocean eddies. Such a model is not possible within the timeframe of this review. The pragmatic solution is to urgently develop and run model resolutions capable of resolving key processes, ie. at the sort of resolution used in current numerical weather prediction (of the order of 50km in the atmosphere). These models will act as stepping stones to higher resolution models.

4. Existing and currently-planned HPC facilities

4.1. Introduction

This section outlines existing and proposes supercomputing capabilities, including planned upgrades and opportunities for collaboration. Facilities are considered against the following two broad categories of activity:

- Production of operational climate science services, including:
    - Operational, often ensemble prediction, systems for forecast lead times of weeks to decades and beyond, including coupled ocean-atmosphere data assimilation to ensure timely, fit-for-purpose services;

  - pre-operational development and testing to pull through new developments into improved operational prediction systems;

- Longer-term research for developing future improved operational systems (typically at least five years hence). Increasingly, these activities will be done in broader collaboration between research institutes, for example NERC and the Met Office, and include:

  - Developing next generation higher resolution modelling capabilities, to explore predictability limits and climate sensitivity, and the use of process-based models for parametrization development:

  - research activities focusing on better representation of weather and climate processes and phenomena, at a level of resolution and sophistication not currently possible with existing models.

4.2. Met Office in-house resource

For many years the Met Office has pioneered a combination of weather forecasting and climate prediction in one organisation and, from the advent of supercomputing technology in the early 1980s, has maintained on its HQ site a single platform for both weather and climate activities. This single platform, coupled with the Met Office's Unified Model (UM), brings major efficiencies which bring added value-for-money in:

- **UM model development** – this benefits both weather and climate applications. For example, work on weather applications is informing the development of next-generation flagship climate models at high resolution. These synergies will increase as data assimilation[4] is included in climate models and as the strategy of delivering a seamless prediction system from days to decades, around a single code base, is implemented;

- **Code optimisation** – timeliness of the production cycle is essential for near-term weather forecasting and the effort put into code optimisation benefits both weather and climate applications;

- **Job scheduling** - the time-critical nature of weather forecasting requires a certain daily level of supercomputing capability at specific times. Certain classes of climate work are run in the schedule gaps when the supercomputer is not required for weather forecasting and thus, for a machine used for both weather and climate work in roughly equal shares, machine efficiency (typically 85%) and hence value for money can be optimised.

A single high-resilience in-house supercomputing capability delivers time-critical weather forecasts 24x7, as well as providing a climate supercomputing service. This

[4] Data assimilation is a term describing any method for combining observations of variables like temperature and atmospheric pressure into numerical models such as the ones used to predict weather.

type of capability will continue to be needed to deliver the following services in a reliable and timely way:

- A growing operational suite of climate prediction runs, with lead times of weeks to decades, including monthly, seasonal, decadal, centennial and Earth System runs;
- Production runs to generate quick, reliable answers to ad hoc requests from Government customers on policy issues (eg. impacts of different emissions policies);
- the 1-2 year ahead research, development and pre-operational testing work which is needed before new prediction systems can be put into production;

Capability upgrade plans: Met Office

The extant Met Office contract with IBM provides for installation of next generation hardware (Power 7) in 2011 which is expected to have a peak performance of 216 TeraFlops for climate purposes and deliver an approximately 3x increase in realisable compute power compared to existing capability. The contract has been flexibly written so that, resource permitting, extra hardware can be provided. The detail of the Phase 2 upgrade will need to be confirmed in November 2010.

One option for meeting increased demand for operational climate services would be to provide more computing resource within the Met Office as part of the existing 24x7 high-resilience capability, or as a lower-cost and lower-resilience variant. Increasing existing Met Office supercomputing resource devoted to climate research by up to 3 x 216 TF peak = 648 TF peak is permissible under the existing contract (no requirement for a new OJEU procurement) and could be accommodated within the existing IT Hall power and cooling infrastructures with some upgrades to power supplies (costing about £3m-6m depending on resilience requirements). Additional running costs associated with this service are estimated at £7m pa.

The contract allows a further increase in capacity to 2 x 648 TF = 1,296 TF peak but would require a new IT Hall to be constructed with a 2 year lead-time and an indicative cost of £20m.

4.3. Met Office / NERC collaboration: MONSooN

NERC-funded researchers have been using the UM as the model of choice for climate research for over a decade. This has mostly entailed using HECToR and predecessor machines. However, in 2009, initiatives to develop closer collaborations resulted in the establishment of a collaborative supercomputer service at the Met Office (called MONSooN[5]) and a commitment to work together to explore re-designing codes for efficient running on future technology architectures. To extend and strengthen collaborative work, the Met Office and NERC launched the Joint Weather

[5] MONSooN: 'Met Office NERC Supercomputing Nodes'.

and Climate Research Programme (JWCRP)[6] in 2009, to grow the UK's leading role in weather and climate research. The JWCRP is an accredited programme of *Living with Environmental Change*[7], comprises some 12 weather and climate joint research projects, and aims to:

- ensure that the UK has access to internationally competitive tools and infrastructure for maintaining its world-leading national capability in observing, understanding, modelling and predicting weather and climate, and their impacts;

- enable closer collaboration between NERC and the Met Office by working to eliminate existing barriers, to align more closely their respective research activities, and to ensure effective participation in relevant new research programmes from both organisations;

- propose new activities to address critical gaps in the existing national portfolio of weather and climate research and to be actively involved in promoting and developing those activities;

- develop mechanisms to promote the more effective pull through of research and development into improved weather and climate forecasts.

Capability upgrade plans: Met Office / NERC

MONSooN will be upgraded to next-generation IBM hardware (Power 7) when the Met Office's own system is upgraded in 2011. At that time, and subject to Met Office / NERC agreement and funding availability, MONSooN could be expanded. This would require upgrades to power supplies as noted above. Additional running costs associated with the service would depend on capacity. Ultimately, the limit on capacity is the same as for the Met Office in-house system. It will be important to ensure that any upgraded service improves on the current provision by providing the full functionality required by the academic climate modelling community (eg. including data transfer services with adequate bandwidth, availability of third-party software, easy access to archived data, and systems for sharing technical information, model code and data).

This system is well-suited to longer-term research and development, and potentially other activities based on collaboration, except regarding its security model. Although this has been relaxed from the operational Met Office environment (and hence is not well-suited to real-time operations) further work is needed to develop a security model which both adequately safeguards the supercomputer and stored data and, at the same time, offers better flexibility and easier access for the academic climate science community.

4.4. National academic supercomputer: HECToR

[6] http://www.metoffice.gov.uk/research/collaboration/jwcrp
[7] LWEC, www.lwec.org.uk

Supercomputing resource for academic research is managed by EPSRC on behalf of three Research Councils (EPSRC, NERC and BBSRC) and is currently provided by HECToR[8]. Currently, the Met Office/MONSooN and HECToR supercomputers use different architectures (IBM and Cray respectively), and so ongoing porting, optimisation and maintenance work is required if models are to be run on both systems.

NERC's 20-30% share of the HECToR resource is mainly devoted to CASCADE and high-resolution climate model development work (eg. HiGEM). This work is increasingly being aligned with JWCRP work run on MONSooN.

Capability upgrade plans: Research Councils

Phase 2b of HECToR is now live and offers about 350 TFlops, of which ~10% on average is currently being used by NERC for climate science. One job could potentially occupy the whole machine, if the capacity was made available.

Phase 3 is due late in 2011 which may entail a 3x increase in the number of nodes, and hence a 3x increase in compute power compared to current, with no change to processor technology.

This system is well suited to longer-term research and development activities, especially those requiring very large resources for short periods. It has excellent network connectivity. The security and access model for this research facility is not well-suited to real-time operations, but is suitable for collaborative work involving the development and provision of climate projections and their interpretation.

4.5. International collaborations around the UM

The Met Office has strengthened its collaborative programmes by sharing the UM forecast and data assimilation system. Links have been forged with operational organisations around the world, by combining skills and capabilities and by aligning research and development programmes[9]. Collaborative research has included model formulation and development (requiring high computational resource with quick turn-around), pre-operational design work and some types of high-resolution production runs (e.g. ensembles).

Met Office groups involved in seasonal to decadal forecasting actively collaborate with academic and operational institutions around the world[10]. Through research and operational licence agreements, recognised centres use the UM for their official duties, such as the provision of public weather and climate services. The Met Office is now collaborating with the Korea Meteorological Administration (KMA[11]) on high-resolution seasonal forecasting with the aim of sharing results from operational seasonal hind-casts and forecasts and effectively increasing the ensemble size, thereby improving confidence in the predictions. Running costs to UK taxpayers are also reduced.

[8] High End Computing Terascale Resources, located in Edinburgh.
[9] http://www.metoffice.gov.uk/research/collaboration
[10] http://www.metoffice.gov.uk/research/areas/seasonal-to-decadal/external-collaborations
[11] http://web.kma.go.kr/eng/index.jsp

KMA's previous optimisation work on related weather models will facilitate optimal running of the corresponding climate models. A similar approach is being discussed with other agencies[12,13].

These opportunities will only be available, however, if the UK can continue to offer collaborating institutions access to world-leading models and science as part of a collaborative and joint operational relationship. The Met Office is seeking a commitment from partners to develop a joint programme to maintain UM capability as world-class whilst reducing user costs. This type of joint activity will be essential for delivering the predictive production capability that will be required in the coming decade.

As collaborations of this type use other agencies' supercomputers. However, exploiting multiple computing locations put stress on existing networks without investment in data archives and light paths to minimise the number of data movements and make them more efficient. This approach (of coordinated data archives and network links) is the basis of the upcoming peta-scale data analysis for the fifth coupled model inter-comparison project (CMIP5), and will appear in the forthcoming European Strategy for Earth System Modelling.

Future opportunities

Where the Met Office/NERC and other (eg. international) agencies have common scientific interests, opportunity exists for initiating further collaborative activities, thus giving UK scientists access to extra results and enabling mutually beneficial knowledge transfer.

However, overseas National Meteorological Services and research institutes will not provide free supercomputing resources to UK scientists and are mostly unlikely to have material spare capacity to sell to the UK. The availability of such resource cannot be guaranteed, and even if it were available costs would be broadly equivalent to the costs of UK provision. It is therefore unsuitable for operational climate services.

4.6. Accessing third party supercomputer resource

Use of remote resources has in the past delivered significant benefit and such opportunities should continue to be explored. However, benefits will need to be set against the extra overhead associated (model validation and integration work) with making the models run effectively on other machines, managing the jobs and handling the output.

Several years ago, the joint NERC/Met Office *UK-Japan Climate Collaboration* gave the UK access to the Japanese Earth Simulator, allowing high-resolution modelling to be started earlier than would otherwise have been the case. Applications were relatively easily transferable, since the Earth Simulator architecture was the same as that used by the Met Office at the time. Development of the coupled climate model was greatly accelerated and the study of tropical cyclones, El Niño-Southern Oscillation (ENSO) variability and many aspects of global-regional climate interaction crucial for climate services

[12] http://www.bom.gov.au/
[13] http://www.ncmrwf.gov.in/

applications, was facilitated. However this required the long-term placement of 4 staff in Japan, and the delivery of model output back to researchers in the UK was extremely challenging.

International developments

The International Centre for Earth Simulation Foundation (ICESF) was established as a response to the World Meteorological Organisation's 2008 *World Modelling Summit for Climate Prediction*[14] and the 2009 *World Climate Conference – 3* (WCC-3[15]). This Foundation seeks to build an international supercomputing facility for Earth simulation, akin to the Japanese Earth Simulator but more internationally available. It would be funded by charitable and industrial donations and would augment national climate centres. The Centre is intended to make significant advances in climate science and modelling through the development of innovative approaches to simulating the multi-scale nature of the Earth system using next generation computing architecture. Funding is not yet committed but it will be important to continue to engage with the Foundation, and assess any opportunities to access new supercomputing resource.

It is expected that, if realised and accessible, such a facility would be most useful for longer-term research or to increase ensemble size in shorter-term research activities.

4.7. EU initiatives

The UK has for many years been involved with various pan-European climate science and model development initiatives (such as ENSEMBLES[16], PRISM and METAFOR[17], and IS-ENES.

The UK should further develop engagement with other European modelling centres to improve seasonal and climate predictions through a multi-model ensemble approach. Collaboration of this type will only be effective if a world-class high resolution UK climate model is available.

PRACE, with funding from the EU's FP7[18], has developed a roadmap for access to, and implementation of, latest supercomputing systems in Europe. Phase 1 (up to 2015) will provide access to 3-5 leading computing systems from mid-2010. Machines will initially be available in Germany (Julich[19], June 2010) and

[14] http://www.wmo.ch/pages/prog/wcrp/documents/WCRP_WorldModellingSummit_Jan2009.pdf This WMO conference considered what advances in climate prediction were needed to address the global climate change challenge and identified (amongst other aims) the need to obtain computational capabilities three to four orders of magnitude greater than the best currently available.
[15] http://www.wmo.int/wcc3/documents/brief_note_en.pdf WCC3 proposed a Global Framework for Climate Services (GFCS) to strengthen the production, availability, delivery and application of science-based climate prediction and services. A task-force has been established to develop the Framework and make proposals on the resource implications of implementation.
[16] http://www.ensembles-eu.org/
[17] http://www.prism.enes.org/
[18] The Partnership for Advanced Computing in Europe (PRACE) prepares the creation of a persistent pan-European HPC service, consisting several tier-0 centres providing European researchers with access to capability computers and forming the top level of the European HPC ecosystem. PRACE is a project funded in part by the EU's 7th Framework Programme.
[19] http://www.fz-juelich.de/jsc/en

France (autumn 2011). Access to PRACE resources will be through peer review according to 'need for access to high-end computing' and 'scientific quality' criteria. PRACE is a legal entity in its own right[20] and will have a head office to run the peer review access system.

Up to 2015, PRACE machines will be provided by the hosting partners (France, Germany, Italy and Spain) and participants will be able to apply to use either part, or all, of the machine for limited periods of time.

> **EU developments: PRACE**
>
> €400m of in-kind funding for the next five years has been committed by France, Germany, Italy and Spain (100MEuro each). Decisions on funding of €100m each from the Netherlands and the UK were delayed due to national elections. RCUK has since confirmed participation, via EPSRC, at a lower of membership level. This will limit the level of access for UK researchers to the PRACE systems. The EC has provided €40m for preparatory and the Implementation Phase 1 projects, and an additional €30m FP7 funding may be allocated to the next Implementation Phase.
>
> Tier-0 computing resources are expected to be available for a range of "grand problems" and not solely for climate research. The architecture of the current Tier-0 machine (an IBM *Blue Gene*) is not ideal for climate science.
>
> Future EU-initiated supercomputing resource is unlikely to be available on a sufficiently-early timescale to allow it to be used to gain better answers to the most pressing climate change questions which HMG is asking. However, the PRACE systems could be used as additional resource for climate change research though code porting and benchmarking would be needed.
>
> After 2014, the next EU Multi-Financial Framework (2014-20) is expected to provide significant additional funding for the climate agenda through the newly-established Directorate General Climate Action[21], which may provide funding for supercomputing infrastructures. Any EU-funded climate supercomputer would only be likely to be operational by around 2017 at the earliest.
>
> As with the proposed ICESF project, it is expected that such a facility would be most useful for longer-term research or to increase ensemble size in shorter-term research activities. Also, for technical work on code scalability, especially where the machines contain novel technologies.

4.8. European Centre for Medium-Range Weather Forecasting (ECMWF)

As a Member State, the UK has had access to a useful supercomputing resource (currently, about 8% of total in-house Met Office resource but reducing to 4% in

[20] Formed as a Belgian association.
[21] http://ec.europa.eu/climateaction/index_en.htm

2011) on ECMWF's machine for some years which has been used for climate uncertainty modelling, seasonal forecasting and THORPEX research[22].

Much of the collaboration with ECMWF is weather-focussed, notably in the areas of data use and data assimilation. However, in light of research showing that better and more reliable seasonal forecasts can be created by combining the output from several models rather than taking just one model, the EUROSIP multi-model seasonal forecasting system[23] has been developed. The Met Office is now working with ECMWF and Météo-France (and in the near future NCEP) to further develop this capability. In the medium-term, this could be a significant contributor to climate services. ECMWF plans to install next-generation IBM hardware (Power 7) in mid-2011. Plans for post-2014 capability are not yet approved by Member States.

Another Met Office / ECMWF joint study, to appraise options for the provision of energy efficient supercomputing services may lead to a joint state-of-the-art green supercomputing service.

4.9. Other opportunities

A multi-agency remote-use team could be established to further exploit future opportunities. Recent examples include:

- partnership of US climate research organisations, ECMWF and two Japanese institutions, funded by the US National Science Foundation, which has investigated the role of model resolution in climate variability and change through simulations using the whole of Oak Ridge National Laboratory's *Athena* supercomputer[24];
- discussions between the National Centre for Atmospheric Research (NCAR) and the Met Office to explore ideas for shared work on the dynamical core of the Unified Model.

However, overseas agencies are mostly unlikely to be interested in collaborating on UK-specific issues, such as development of UK regional climate models, and so cannot be relied upon for any critical UK climate programme work. Staff costs associated with use of opportunistic supercomputing services can also be high.

5. Supercomputing capabilities: delivery and sizing options appraisal

In developing the options included in this report, it is assumed that the scientific capability of the Met Office Hadley Centre and NERC-funded climate activities will be maintained. **There are strong economic and security arguments for HMG**

---

[22] THORPEX http://www.wmo.int/pages/prog/arep/wwrp/new/thorpex_new.html is a WMO 10-year international research and development programme to accelerate improvements in the accuracy of one-day to two-week high impact weather forecasts for the benefit of society, the economy and the environment.

[23] This consists of three independent coupled systems: ECMWF, Met Office and Météo-France, all integrated in a common framework. The reason for creating a multi-model forecasting system is that in most cases, the multi-model combination is better than the best single model. http://www.ecmwf.int/products/forecasts/seasonal/documentation/eurosip/ch1.html

[24] http://www.ornl.gov/info/press_releases/get_press_release.cfm?ReleaseNumber=mr20091116-01

continuing to have a centre of excellence in climate change prediction and service provision under its control, so that it has access to the best possible science and advice in a timely and secure manner. This also ensures that there is consistency of approach across HMG and avoids fragmentation in the delivery of both the science and the predictions.

This assumption recognises that the:

- UK's unique world-leading science capability, built upon the synergies between joint weather forecasting and climate modelling activities, and collaborative work such as the JWCRP, cannot currently be matched by other countries' capabilities. Three UK institutes, and a European one based in the UK, are listed in the top ten of the world's leading geosciences research centre[25], with the Met Office Hadley Centre as number one;

- UK-based scientific community will need to be maintained if the urgent climate science questions are to be answered to the level of rigour and on the timescales required by HMG. The option of relying on third party scientific advice and investing no further in UK supercomputing capability is considered overly-risky as there would be no guarantee that advice would be available to answer HMG's particular questions. In particular, overseas institutions are unlikely to focus on the higher-resolution regional modelling that is needed to assess UK climate impacts.

In assessing options the following factors also need to be considered where different levels of engagement are appraised:

- **Climate collaborations, especially around the UM** – noting the recently created JWCRP and associated MONSooN capability, and the increasing collaborative take-up of the UM overseas, it will be important to build on these initiatives to speed scientific advance. Opportunities undoubtedly exist for the Met Office to collaborate more-closely with the UM partners on activities such as ensemble generation (thereby potentially offsetting some HMG supercomputing costs) and to work more-closely with NERC and other partners on model development using shared supercomputing resource, provided suitable arrangements are made regarding joint development of the UM, sharing of code and other technical information. Joint research programmes are likely to provide the most-suitable governance framework for this work;

- **European climate supercomputing provision** – it is unlikely that a pan-European supercomputing resource, suitable for the type of climate work needed by the UK, will be available for some 5-8 years. The UK's current level of engagement with EU initiatives is modest (eg. contributing to the PRACE partnership at the lower membership level. A more pro-active UK stance could help ensure that EU supercomputing provision is more-closely aligned to UK interests. Further, the UK could position itself to become a major hub on any future EU climate supercomputing network, by building on its world-leading climate science position. Longer-term opportunities for

[25] http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=409181&c=1

switching HMG funding from UK to EU-based facilities could then be reviewed if it were shown that this would offer better value for money.

Summary results of the options appraisal are shown in Table 6 below:

	Appraisal dimension	**Units**	**Option**				
			A	**B**	**B+**	**C**	**D**
1	Indicative cost of service	£m pa	3	6	9	26	59
2	Does option allow harnessing of available science?	Yes / No	Yes	Yes	Yes	Yes	Yes
3	Are codes sufficiently scalable to be used immediately?	Yes / No	Yes	Yes	Yes	Yes	No
4	Is IT Hall accommodation immediately available? (if not, approx. lead time)	months	Yes	Yes	Yes	Yes	No (9-12 mths[26])
5	Are HMG's questions answered to required timescale?	Yes / No	No	No	Partly (5/8 questions answered)	Yes	Yes

Table 6: Summarised options appraisal results

Planned Joint Code Development Work with the Hartree Centre

NERC has committed £2.2m over 5 years, with matching funding in kind from the Met Office and a similar level of support expected from STFC via the proposed Hartree Centre[27], "to research, design and develop a new *dynamical core* suitable for atmospheric modelling on massively parallel computers, and to produce a demonstration integration that provides scientific insight into the impact of hitherto unresolved processes on global atmospheric dynamics". The inclusion of Hartree will provide access to a range of different architectures, so any code developed will be as future-proofed as possible against changes in technology architecture.

How much compute power is needed

Table 7 shows the relative cost of climate models that the Met Office expects to use over the next 15 years, assuming the same length of integration and number of ensemble members or scenarios that need to be run. The final column shows how long it would take to provide this capacity assuming Moore's Law improvements only.

Model Ref name	**Climate model name**	**Horizontal resolution and number of vertical levels**		**Cost relative to Model A**	**Moore's law years**
		Atmosphere	Ocean		
A	HadGEM2-ES	135km, 38L	72km, 40L	1	0
B	HadGEM3-ES	135km, 85L	72km, 75L	6	4
C	HadGEM3	135km, 200L	17km, 75L	115	11
D	HadGEM3	60km, 85L	17km, 75L	18	7
E	HadGEM3-ES	40km, 85L	6km, 75L	489	14

[26] Using current modular construction and installation engineering methods.
[27] In particular through their topic Computational Challenges in the Environmental Sciences, see http://www.cse.scitech.ac.uk/events/Hartree_Summary/

F	HadGEM3-ESv2	25km, 85L	6km, 75L	1463	16
G	**HadGEM3-ESv2**	**1.5km, 200L**	**6km, 75L**	**406123**	**28**
H	UK HadGAM3	1.5km, 70L covering the UK domain	N/A	227	12

Table 7: Relative costs of various climate models

Model A has been used as the baseline as it is affordable at all timescales and is currently used for centennial runs and scenario and some ensemble work. More expensive models are used for development on next generation models and for shorter runs, such as seasonal runs, and cheaper models are used for some ensemble work. It is not known when a 1km resolution global model (Model G) would be sufficiently scalable to run at useful speeds for climate research because changes in technology and options for parallelism are too far away from current experience. However, this is unlikely to be practical in the next 15 years. Models E and F are practical within that timescale, but would not be usable for any sorts of climate runs until about 2016 and more likely post 2020 for longer runs.

As well as computing capacity, the science needs to be developed and the model performance needs to be improved to allow enough scalability to turn capacity into performance. Table 8 below shows when the science and scalability will be ready:

Question	**Answer question by**	**Example model required in answering the question**	**Limitations**	
			Science not ready until	**Model scalability not ready until**
Q1	ASAP	D: HadGEM3 A(60km, 85L) O(17km, 75L)	2010-2013	2011
Q2	2011-16	B: HadGEM3-ES A(135km, 85L) O(72km, 75L)	2012	Now
Q3	2016	H: UK HadGAM3 A(1.5km 70L)	2013	2012
Q4	2015	C: HadGEM3 A(135km, 200L) O(17km, 75L)	2015-2020	2016[28]
Q5	2015	B: HadGEM3-ES A(135km, 85L) O(72km, 75L)	2015-2020	Now
Q6	2020	E: HadGEM3-ES A(40km, 85L) O(6km, 75L)	2015	2016[29]
Q7	2020	F: HadGEM3- ESv2 A(25km, 85L) O(6km, 75L)	2018	2018-2022[30]
Q8	2020-25	H: Large domain	2018	2022

[28] It is hard to scale to increased vertical level s in the current UM structure because it does not support vertical parallelism. More level sand shorter timestep will demand scaling.
[29] Will require a new scalable coupler and more ocean scalability.
[30] Will need to wait on the new dynamical core to be proven in a climate model.

		HadGAM3 A(1.5km 70L)		

Table 8: Dates by which science and scalable code will be ready

For each question one can make a quantitative estimate of the amount of compute power needed ("model cost factor" in the following table), considering things like run length and number of runs. Based on Moore's Law, and assuming a 2 year upgrade cycle (as used at ECMWF), one can calculate a funding ratio to provide the capacity required to answer the question on the listed date in Table 9 below:

Question	Example model required in answering the question	Date to answer	Model cost factor[31]	Level funding HCP Capacity	Funding ratio to answer[32]
Q1	D: HadGEM3 A(60km, 85L) O(17km, 75L)	2012[33]	4	3.0	1.5
Q2	B: HadGEM3-ES A(135km, 85L) O(72km, 75L)	2012	12	3.0	4.0
Q3	H: UK HadGAM3 A(1.5km 70L)	2016	5	19.0	0.25
Q4	C: HadGEM3 A(135km, 200L) O(17km, 75L)	2015	38	7.6	5.1
Q5	B: HadGEM3-ES A(135km, 85L) O(72km, 75L)	2015	15	7.6	1.9
Q6	E: HadGEM3-ES A(40km, 85L) O(6km, 75L)	2020	92	121	0.76
Q7	F: HadGEM3- ESv2 A(25km, 85L) O(6km, 75L)	2020	366	121	3.0
Q8	H: Large domain HadGAM3 A(1.5km 70L)	2022[34]	1135	355	3.2

Table 9: Capacity required

What this analysis does not do is look at the full range of activities required, such as model development, to answer key questions. However, as the baseline for the analysis has been the current, most commonly used model, it is reasonable to assume that the range of current activities can be similarly scaled to fit around these activities. But it can be seen that several core activities happen concurrently. Table 10 estimates this effect:

Year	Activities	Aggregate funding estimate
2012	Q1 + Q2	5.0
2015	Q4 + Q5	5.0[35]
2016	Q3 + residual Q4 + Q5	2.3[36]
2020	Q6 + Q7	3.8
2022	Q8	3.2

[31] Includes a consideration of run length and number of runs (scenarios, ensembles, etc).
[32] Approximate ratio of increased cost of model to increase in compute power delivered by Moore's law
[33] ASAP=First available date because of HPC upgrade dates
[34] 2020-2025 assumed as 2022
[35] Note that Q4 is not scientifically ready until 2016. Here we have assumed it starts in part in 2015 ahead of current planning.
[36] 2016 sees an upgrade, so Q4 now only requires a 2x funding ratio

Table 10: Impact of concurrent activities

Performance and capacity analysis

As Figure 2 below shows, for a given model, the speedup provided by adding more compute resource to the problem tails off. Pushing models to get higher performance will, increasingly, reduce the efficiency of those models and hence reduce the number of model years-per-day that can be achieved on the whole resource.

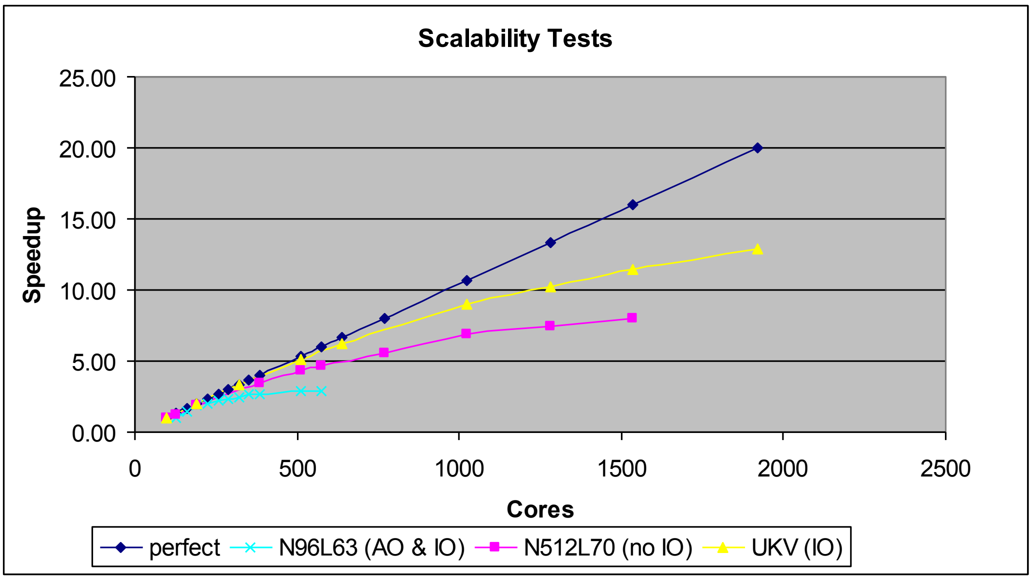

Figure 2: Performance vs compute resource

Increasing computing capacity to gain increased performance gives a sub-linear improvement and, as can be seen in the graph, eventually gives very little improvement.

As models increase in resolution, they can scale out to use more processing elements (N512 scales better than N96 in the graph above). Nevertheless, scalability will be increasingly challenging because the calculations per horizontal grid point increase due to shorter time-steps, more vertical levels and more scientific complexity of future models.

Historically, CPU performance has increased over time, aided by modest increases in parallelism. This has met the scalability challenges to date but will no longer help as new computers are made more powerful by having more processing elements rather than faster processors. This trend helps with capacity, but not performance.

Options to improve performance are:

- The purchase of high performance systems. These have better processor performance, better memory bandwidth to use that processor performance and better performing networks to reduce the overheads of parallelism;

- General, machine specific, optimisations to make best use of processing capacity to limited the required scalability;

- Increasing code parallelism. E.g. running the IO systems in parallel with the science, running model components in parallel (Atmosphere, Ocean, Chemistry), running the physics in parallel. The latter two options would imply scientific compromises which would take time to develop and implement;

- Develop new, more scalable, algorithms, such as a new dynamical core.

Hence, performance issues can stop certain models being run at a given time, even if capacity is available. For example, an analysis of high resolution climate models (based on a 25km atmosphere) that are capable of scaling to run at 2 model years per day suggests that this is not possible until we have implemented a new grid and find new opportunities for increased parallelism, both of which will have scientific impact that will need to be validated and tuned. The analysis shows that it would only be possible to do limited, shorter runs, at high cost and less than 2 years per day by 2015 even with a trebling of funding, assuming the model could be developed and validated on that timescale.

A demanding performance requirement will need to be met by a range of measures, including high performance systems and significant model changes. The need to explore the full range of measures will increase as the challenges increase. The challenges of the performance requirement will set dates before which given resolutions are not practical for climate research activities, for example, significant use of a 25km model is not practical before 2018 for technical rather than scientific or capacity reasons.

31 August 2010

This publication is also available on our website at http://www.bis.gov.uk
Any enquiries regarding this publication should be sent to:

Department for Business, Innovation and Skills
1 Victoria Street
London SW1H 0ET
Tel: 020 7215 5000

If you require this publication in an alternative format, email enquiries@bis.gsi.gov.uk, or call 020 7215 5000.

URN 10/1290AN